Information and Knowledge Management
ISSN 2224-5758 (Paper) ISSN 2224-896X (Online)
Vol.7, No.10, 2017

Big Data Attributes and Knowledge Discovery Process: An Empirical Analysis of the Anticipated Mediating Role of Cloud Computing

Safinaz BanyFayyad
Independent Researcher, Amman, Jordan

Dr. Firas M. Alkhaldi*
Department of MIS, Faculty of Business, Amman Arab University, Amman Jordan

Abstract
This study attempt to investigate whether cloud computing can act as a facilitating agent in support of knowledge discovery from big data. The study proposed a number of propositions to investigate this possible effect and impact between computing cloud, big data and knowledge discovery. The telecommunication industry was selected as the research population and the study sample covered the main leading telecommunication companies in Jordan. A survey questionnaire was developed and distributed to a selected study sample. The proposed models of the study were tested using factor analysis and PLS method. The results indicated that there is no mediation effect as proposed by the research model between cloud computing characteristics and knowledge discovery processes via could computing. The results also revealed that big data attributes has a direct significate impact on a selected knowledge discovery processes and a selected cloud computing characteristics. The study also advised on some interesting finding on the three domains of the study.
Keywords: Cloud computing characteristics, big data attributes, knowledge discovery processes in database, PLS method.

1.0 Introduction

Continues changes and technological developments in various fields such as social, political, economic, technological, requires the presence of techniques and tools to cope with and to adjust to these changes, specifically in the field of technology. Due to the rapid changes that is occurring, there must be a capable and effective tools to help in adapting to these changes. Cloud computing, Big Data, and Knowledge Discovery from Databases were emerged as one of the most important tools and services in response to these changes, the study will look into the capacity of this tool with regard to the concept of big data and knowledge discovery. In recent years cloud computing received much considerable attention by academics and professionals, to the point it had become a sort of technological revolution. Cloud Computing is a term refers to set of integrated services, techniques, applications and systems that serve the customer remotely without a necessarily need of local resources, with attractive characteristics which make it desirable and acceptable. This research attempt to investigate the possible mediating relationship that this concept might play in support of knowledge discovery from big data, as big data has become a phenomenon in various and large-size and medium-size enterprises and companies, so it was necessary to find tools and techniques to deal with this huge amount of data, to extract what is essential to accomplish the required transactions and facilitate the handling of data, to enhance the organization competitive standing. As far as knowledge discovery, the concept attracted many of the research articles and industries, and it and had become a focus of much attention (Fayyad, et al., 1996).

2.0 Cloud Computing (CC)

Armbrust, et al. (2010) indicated that CC refers to software, hardware, and capabilities delivered as a service over the web with the support of various technologies such as advance technologies in networks and data centres, these services are classify based on the nature of service; Platform, Infrastructure and Software as a Service. Jones (2009) also indicated that cloud computing is simply the migration of computing and storage outside an enterprise and into the cloud, Where the user define the needs of the computing resources , then Cloud Computing service provider compile these requests and submit them to the customer. Cloud Computing is a set of integrated services, techniques, applications and systems that serve the customer remotely without a necessarily need of local resources, with attractive characteristics which make it desirable and acceptable such as resource pooling, Broad network access, Virtualization, security, Rapid elasticity, scalability, and on-demand service.

2.1 Cloud computing characteristics

There are several characteristics of cloud computing which make it attractive to be adopted, as following:
Shared resource pooling: where the infrastructure's provider suggests set of processing technologies that can be

gradually doled out to many technology customers. Such exclusive technologies task oriented capability provides much flexibility to base assets providers helping in their specific efforts of asset utilization and operations expenses (Zhang, et al., 2010). It is also Sharing of "resources and costs" among large group of users allows efficient utilization of the infrastructure (Jadeja & Modi, 2012). Resource pooling implies a multi-residency, for example, a physical server may have a couple of virtual machines having a place with unmistakable customers (Puthal, et al., 2015). Resource pooling allows combining resources of computing (e.g. hardware, software, processing, network bandwidth) to provide various consumers (Mahmood, 2011). It can be seen as cloud computing provider allows users to share a various virtual and physical resources as they allocated and reallocated giving the needs of a consumer (Bhardwaj et. al., 2010; Mell & Grance, 2011). Durkee (2010) indicated that resource pooling as to Cloud computing services allow customers layers infrastructure participation, in order to save the consumption of the size of the network.

2.1.1 Broad network access

Clouds are generally accessible through the internet and use the internet as a service delivery network. Hence any device with internet connectivity, such as a mobile phone, a PDA or a laptop, is able to access cloud services (Zhang, et al., 2010). Broad network access reflects the competencies that exists within any computer network and retrieved thru a regular technologies that uses either thick and/or thin client platforms as in PDAs, PC, cellphone, laptop, and workstation (Mell & Grance, 2011). In cloud computing, clients get to the information, applications or whatever other administrations with the assistance of a program paying little attention to the gadget utilized and the client's area.

2.1.2 Service oriented

Zhang, et al., (2010) pointed out that as in any technology service contact as well as in cloud computing, the service contracted is always govern by a Service Level Agreement (SLA). Well clear thought and convenience, accessibility are the core answers to attain service oriented notion. In that sense services such as virtualization and other advanced technologies will ease service with the burdens of exposing much efforts to user (Gong, et al., 2010).

2.1.3 Virtualization

Cloud assets are frequently virtualized as a service over the Internet. Server farm utilizes virtualization advances which theoretical the ordinariness of foundation in various levels (Gong, et al., 2010). Virtualization is to utilize equipment or programming to make the perception of something. Virtualization is a critical empowering innovation that digests framework and assets to be made accessible to customers as segregated VMs (Takabi & Joshi, 2010).

2.1.4 Security

Security is one of the most critical characteristics of cloud computing that make clients adopt this technique to perform their transactions (Zhang, et al., 2010). High security of CC is accomplished for the most part through three ways. Firstly; the free coupling makes cloud computing framework run well when some portion of it is demolished. Second, the deliberation, virtualization and privation of cloud supplier abstain from uncovering the points of interest of relating usage. Third, innovation participating with law is the watch of distributed computing (Gong, et al., 2010). Security can be the same as or superior to conventional frameworks since suppliers can dedicate assets to tackling security issues that numerous clients can't manage (Jadeja & Modi, 2012).

2.1.5 Scalability and elasticity

Alkhaldi and Qararah (2016), asserted that cloud computing will support and offer a high range of scalability and elasticity for end users and organizations where scalability means that performance can be monitor and thus it is scalable (Jadeja & Modi, 2012). Rapid elasticity and scalability allows functionalities and resources to be rapidly and automatically provisioned and scaled (Mahmood, 2011). Users swiftly can increase resources from cloud by reducing out and roll back by discharging those resources once they are no more needed (Puthal, et al., 2015). Rapid elasticity and scalability allows functionalities and resources to be rapidly and automatically provisioned and scaled (Mahmood, 2011). Rapid elasticity is Capacities that can be flexibly provisioned and discharged, now and again naturally, proportional quickly outward and internal proportionate with request. To the buyer, the capacities accessible for provisioning regularly give off an impression of being boundless and can be appropriated in any amount whenever (Mell & Grance, 2011). Rapid elasticity is cloud computing capability to increase and decrease the capacity according to customer demand (Bhardwaj, Jain, & Jain, 2010). Elasticity indicates that Computing is provided in the amount required and disposed of when no longer needed (Durkee, 2010)

2.1.6 On-demand self-service

Cloud computing offers fast fulfilment of demand for computing and has a continuing ability to accomplish such demand as required (Durkee, 2010). ODSS indicates that service buyer can individually manage and govern its utilization of computing capacities and capabilities as required autonomously without any necessary interfacing with the designated service supplier (Mell & Grance, 2011).

In order to assess the most important characteristics of cloud computing, the researchers reviewed the many

previous studies relevant to this area and revealed the most important characteristics mentioned in those studies, and accordingly, the following table shows the most mentioned characteristics , The most Seven ones (Resource pooling, Broad network access, Virtualization, Security. On-demand service, Elasticity and Scalability).

Table 1. Cloud computing characteristics as reported by various studies.

Characteristics	References
Shared Resource Pooling	Zhang, et al., (2010); Jadeja & Modi (2012); Puthal, et al. (2015); Mahmood (2011); Mell & Grance (2011); Bhardwaj, et. Al., (2010); Durkee (2010) Takabi & Joshi 2010; Bezemer & Zaidman (2010)
Broad Network Access	Alkhaldi and Qararah (2016); Zhang, et. al., (2010); Jadeja & Modi (2012); Puthal et. al., (2015); Bhardwaj et.al., (2010); Mell & Grance (2011)
On demand Service Oriented	Zhang, et. al., (2010); Gong et. al., (2010); Puthal, et. al., (2015); Mahmood (2011); Bhardwaj et. al., (2010); Mell & Grance (2011); Durkee (2010); Takabi & Joshi (2010)
Virtualization	Gong & et. al., (2010); Puthal et. al., (2015); Takabi & Joshi (2010);
Security	Alkhaldi and Qararah (2016); Zhang, et. al., (2010); Gong et. al., (2010); Jadeja & Modi (2012)
Scalability and elasticity	Alkhaldi and Qararah (2016) Jadeja & Modi (2012); Puthal et. al., (2015); Mahmood (2011)

3.0 Big Data

Big data phenomenon refers to the practice of collection and processing of very large data sets and associated systems and algorithms used to analyze these massive datasets (Begoli & Horey, 2012). The term used to reflect on the expansion in the mass and capacity of data that is difficult to store, organise, and analysis over traditional database. The expression "Big Data" is moderately new in IT and business (Abaker, et al., 2014). Most of the Big Data surge is data in the wild unruly substance like words, images and video on the web and those streams of sensor data. It is called unstructured data and is not typically directed for traditional databases (Lohr, 2012). Due to the complexity of big data, innovative architectures and advanced technologies is essential as a tools and methodologies to be able to generate values form big data (Katal, et al., 2013).

3.1 Big data attributes: The three VS

The 3Vs (volume, variety and velocity) are three defining properties or dimensions of the concepts of big data.

3.1.1 Volume

Refers to the large amount of data, where the amount of data to analyze is enormous. Formerly, data created by conventional means and mainly contributed to employees, currently the data are generated by machines, technology networks and people's interaction in systems such as social networks. Focuses on the quantity of data (Brauna, et al., 2016). The main distinctive attribute that makes the data "big" is the sheer volume. It refers to the constant and dynamic measure of a various set of data produced from several sources. (Abaker, et al., 2014). Volume means with the creation and gathering of masses of data, data scale becomes increasingly big (Chen, et al., 2014).

3.1.2 Velocity

For some applications, the speed of data creation is significantly more imperative than the volume. Velocity specifies the speed of data exchange. Velocity implies the convenience of huge data, particularly; data gathering and examination, and so on must be quickly and convenient led, in order to maximally use the business estimation of enormous data (Chen, et al., 2014). Velocity in context of big data is the understanding and managing the speed of the data initiated from various sources and from which the data streams (Katal, et al., 2013). Velocity focuses on the data collected or generated speed (Brauna, et al., 2016).

3.1.3 Variety

It demonstrates the different sorts of data, which incorporate semi-structured and unstructured data, for example, sound, video, site page, and content, and in addition conventional organized data (Chen, et al., 2014). Katal, et al., (2013) stated that variety reflects the possible multiple classification as it includes the various data structured aggregated from many sources, such as emails, web files, social networking, data sensor generated, etc.... Variety focuses on differences in types, contents, or formats of data (Brauna, et al., 2016).

Many big data attributes were cited in related literature, the researchers' strategy was to adopt and select only the most significant and the most frequently cited attributes of the big data, attributes selected are shown in the following table.

Table 2. Big Data attributes as reported by various studies

Attribute	Reference
Volume : (BDVOL)	Begoli & Horey (2012); Abaker et. al., (2014); Brauna et. al., (2016); Chen et. al., (2014); Katal et. al., (2013); Chihoub, (2013); McAfee & Brynjolfsson (2012).
Velocity : (BDVEL)	Abaker et. al., (2014); Brauna et. al., (2016); Chen et. al., (2014); Katal et. al., (2013); McAfee & Brynjolfsson (2012).
Varity : (BDVAR)	Abaker et. al., (2014); Brauna et. al., (2016); Chen et. al., (2014); Katal et. al., (2013); McAfee & Brynjolfsson (2012).

4.0 Knowledge Discovery

This section presents a general overview of Knowledge discovery, starting with the concept of knowledge and moving to knowledge management finishing with knowledge discovery process.

4.1 Knowledge discovery in databases (KDD)

KDD is a programmed, exploratory examination and displaying of vast databases. KDD is the composed procedure of distinguishing legitimate, novel, helpful, and justifiable examples from extensive and complex database collections (Maimon & Rokach, 2010). KDD is the nontrivial procedure of recognizing appropriate, unique, conceivably significant, and inevitably justifiable examples in data. KDD procedure includes utilizing the database alongside any imperative choice, preprocessing, sub inspecting, and changes of it; applying data mining techniques (algorithms) to organize designs from it; and evaluating the aftereffect of data mining to distinguish the subset of the specified examples esteemed knowledge (Fayyad,et al., 1996). Knowledge Discovery in Databases (KDD) is new paradigm that spotlights on discovering a lot of data and on revelation of significant and intriguing examples inside them. While most work on KDD is worried with organized databases, unmistakably this worldview is required for taking care of the gigantic measure of data that is accessible just in unstructured literary shape (Feldman & Dagan, 1995).

The knowledge discovery from real-life databases is a multi-phase process consisting of numerous steps, including attribute selection, disconnection of real valued attributes, and rule induction (Zhong & Skowron, 2001). The purpose of KDD is to present “valuable” information to the user (Feldman & Dagan, 1995). Knowledge Discovery from Data (KDD) refers to a set of activities designed to extract new knowledge from complex datasets (Begoli & Horey, 2012).

4.2 Knowledge discovery processes

The KDD procedure includes utilizing the database alongside any required choice, preprocessing, sub examining, and changes of it; applying information mining strategies (algorithms) to list designs from it; and assessing the results of data mining to distinguish the subset of the specified examples considered knowledge (Fayyad et al., 1996). Researchers and scientists indicated that knowledge discovery processes included five steps as following:

4.2.1 Selection

is the primary step is creating an understanding of the application dimensions such as space, gathering important information, and differentiating client's information mining objective from others. So, the objective information utilized as a part of the knowledge revelation process can be picked, i.e., choosing appropriate information tests and an important subset of factors (Ristoski & Paulheim, 2015). Selection essential objective is to make an objective data index from the original data, i.e., choosing a sample of elements or information assessments, on which disclosure must be achieved (Gullo, 2015).

4.2.2 Preprocessing

Data is prepared in a way that takes into account a consequent investigation. Preprocessing intends to "clean" data by performing different operations, for example commotion demonstrating and expulsion, characterizing legitimate techniques for dealing with missing data fields, representing time-succession data (Gullo, 2015).

4.2.3 Transformation

Is the third step aimed to create a projection of the data to a frame that data mining algorithms can take a shot at—as a rule, this implies transforming the data into a propositional shape, where each occasion is spoken to by a component vector. To enhance the execution of resulting data mining algorithms, dimensionality diminishment techniques can likewise be connected in this progression to lessen the compelling number of factors under thought (Ristoski & Paulheim, 2015). Transformation is responsible for decreasing and projection the data, so as to infer a portrayal appropriate for the particular assignment to be performed; it is normally refined by including change procedures or techniques that can discover invariant portrayals of the data (Gullo, 2015).

4.2.4 Data mining

Once the data is available in a valuable organization, the underlying objective of the procedure is coordinated to a specific technique, for example, order, relapse, or bunching. This advancement integrates the process of selecting which models and parameters, and managing a precise data mining scheme with the overall criteria of

the KDD practise (Ristoski & Paulheim, 2015). Data mining manages extracting valuable models by picking a particular data mining technique or task, in addition to pick legitimate algorithm(s) for performing out the job that needs to be done, and a fitting portrayal of the yield comes about (Gullo, 2015).

4.2.5 Evaluation and interpretation

In the last step, models determined by the data mining calculations are inspected as for their legitimacy. Besides, the client evaluates the handiness of the discovered information for the given application. This development can equally embrace the views of the distinct illustrations and models, or interpretation of the data exploiting the extracted models (Ristoski & Paulheim, 2015). Interpretation/evaluation is abused by the client to decipher and extricate knowledge from the mined examples, by picturing the examples; this translation is ordinarily completed by imagining the examples, the models, or the data given such models and, on the off chance that, iteratively glancing back at the past strides of the procedure (Gullo, 2015).

5.0 Research Model and Hypotheses

This study was carried out in order to investigate the impact of big data attributes on knowledge discovery processes and the possible mediation role of the cloud computing characteristics in facilitating such impact. The study selected the telecommunications industry as the study population which is known as the most advanced technology oriented companies in Jordan. The study adopts a descriptive and analytical approaches in order to interpret the relationships between study variables. This approach describes the phenomenon of the study in order to reveal the occurrence causes and factors and control them and then extract the results and generalize it. The study sample of the communication population is selected as a non-random purposeful sample which consist of the three main telecommunication solutions provider and the most technologically advanced one (Zain, Orange, and Umniah). As for research respondents, they were selected by a contact person in each company based on their roles, capabilities, and knowledge that fits the research knowledge criteria. The hypothesed model was developed (see figure 1) to test the intended relationship and the possible impact of as indicated earlier. The research "hypothesed" model consist of three measurement models; Big Data Attributes (BDA), Knowledge Discovery in Database Processes (KDDP), and Cloud Computing Characteristics (CCC). Each sub model will be tested and verified against its elements (see table 3) using explanatory and confirmatory factor analysis, and based on findings the Sub-models and the overall structural research model will be modified if necessary and tested against the research hypotheses.

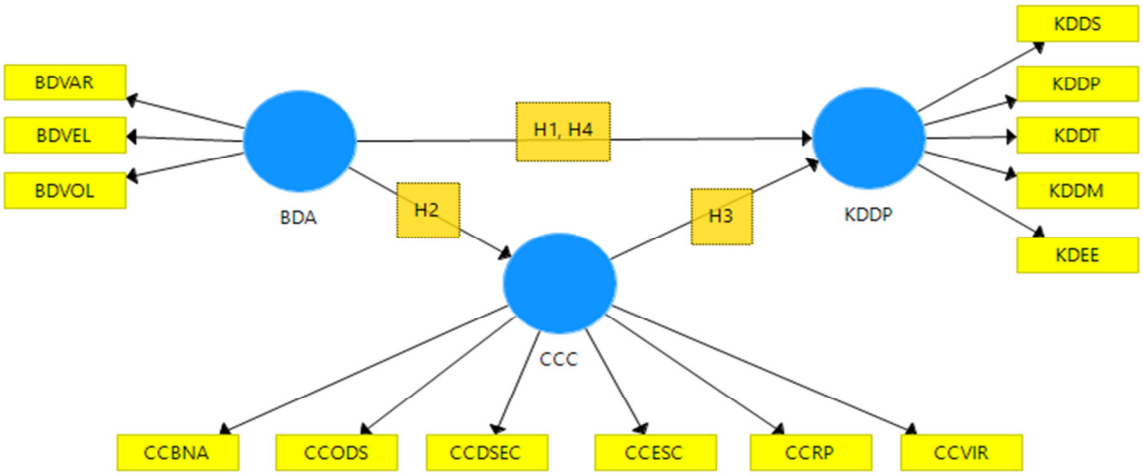

Figure 1. Hypothesized Research Model

Table 3. Research model's element description

Contract: Measurement Model	Variables	Code
Big Data Attributes : BDA	Volume	BDVOL
	Velocity	BDVEL
	Varity	BDVAR
Knowledge Discovery in Database Processes: KDDP	Selection	KDDS
	Pre-processing	KDDP
	Transformation	KDDT
	Data mining	KDDM
	Evaluation and Interpretation	KDEE
Cloud Computing Characteristics: CCC	Resource pooling	CCRP
	Broad network access	CCBNA
	Virtualization	CCVIR
	Security	CCSEC
	Rapid elasticity Scalability	CCESC
	On-demand service	CCODS

5.1 Research Hypotheses

H1: There is a significant impact of big data attributes on knowledge discovery processes
H2: There is a significant impact of big data attributes on Cloud Computing Characteristics
H3: There is a significant impact of Cloud Computing Characteristics on knowledge discovery processes
H4: There is a significant mediation effect of big data attributes on knowledge discovery processes via Cloud Computing Characteristics

6.0 Data Analysis

Data were collected from the three major Jordanian telecommunication companies were distributed to a total of 100 respondents, of which 88 completed questionnaires were retained and analyzed. The data was test for normality using the common indicator (SD, Skewness and kurtosis) the results as in table 4 revels that the sample data was normally distributed and thus it's valid for further testing and data analysis process.

Table 4. Descriptive statistics and normality test

Construct	Variables	Code	Mean	Std. Dev	Skewness	Kurtosis
Cloud Computing Characteristics	Broad Network Access	CCBNA	4.47	.499	-.593	-.703
	Resource Pooling	CCRP	4.23	.514	-.489	-.919
	Virtualization	CCVIR	4.34	.4834	-.252	-.585
	Scalability	CCSCA	4.36	.513	-.735	.451
	Service Oriented	CCSO	4.37	.462	-.686	-.182
	Security	CCSEC	4.28	.335	.119	-.794
Big Data Attributes	Volume-Ownership	BDVOO	4.13	.601	-.096	-.855
	Volume-Operation	BDVOP	4.35	.558	-.607	-.073
	Varity	BDVAR	4.28	.515	-.977	.497
	Velocity	BDVEL	4.34	.511	-.707	.361
Knowledge Discovery process	Data Selection	KDDS	4.42	.513	-.492	-.700
	Data Processing	KDDP	4.25	.496	-1.144	.992
	Transformation	KDDT	4.21	.598	-.638	-.354
	Data mining	KDDM	4.28	.556	-.282	-.693
	Evaluation and Interpretation	KDEE	4.24	.485	-.384	-.991

6.1 Sample characteristics

The average respondents' age falls between 36-35, of which 48.6% are between the ages of 26-30, and 85% are between the ages of 26-35. For the respondent's level of education; the most frequent number of level of education is Bachelor's degree with 52% and the next frequent number is postgraduate degree with 43%. For gender 53% of sample was female.

6.2 Explanatory factory analysis

The proposed model was analyzed using explanatory factor analysis using SPSS V23. All tested factors showed one factor solution except for big data volume, were the results suggested two sub factors for big data volume; one reflecting volume as a results of ownership (BDVOL-O), and the other suggested volume as a results of operation (BDVOL-OP). The overall results as displayed in table 5. Indicators included are; Kaiser-Meyer-

Olkin (KMO), which is a measure of how suited the data is for factor, the results showed that KMO in most case did exceed the accepted benchmark of 0.50, while factor loading in both dimensions was above the acceptance criteria of 0.50 (Hair et, al., 2010). The total variance explained (TVE) show a strong explanation power for all items ranging from 33% to 73%.

Table 5. Explanatory factor analysis results

Construct	Variable	KMO*	Loading Range	TVE**
CCVCs	CCBNA	0.50	0.836-0.972	70.41%
	CCSRP	0.56	0.505-0.786	47.83%
	CCVIR	0.43	0.616-0.907	56.47%
	CCESC	0.50	0,662-0.790	0.62.44%
	CCSO	0.50	0.853-.0856	73.87%
	CCSEC	0.50	0.712-0.712	51.98%
BDAs	DBVOL_OO	0.52	0.816-0.915	38.79% (71.89%)
	DBVOL_OP	0.52	0.713-0.973	33.09% (71.89%)
	DBVAR	0.64	0.740-0.821	59.41%
	BDVEL	0.62	0.600-0.825	55,26%
Knowledge Discovery process	KDDS	0.65	0.755-0.800	60%
	KDDP	0.50	0.552-0.846	51%
	KDDT	0.62	0.724-0.859	62%
	KDDM	0.50	0.801-0.809	66%
	KDEE	0.58	0.602-0.776	50%

* Kaiser-Meyer-Olkin Measure of Sampling Adequacy. **Total variance explained

6.3 Measurement model quality criteria

As for the quality criteria of the sample data; Cronbach's alpha and rho_A measures was used as an indicators of internal consistency of the developed scale, where the average value extracted (AVE) and the composite reliability was used to establish validity of the scale. The results showed an accepted level on consistency as both measures did fall within the acceptable common benchmark, and an acceptable level of validity of the scale was established as CR is within common benchmark, while AVE fall little short of the accepted domain, yet the validity results should deem acceptable, as CR is exceeding the 0.7 mark, see table 6.

Table 6. Quality criteria: reliability and validity tests

Construct	Reliability test		Validity tests	
	Cronbach alfa	Rho_A	Composite reliability (CR)	The average variance extracted (AVE)
BDA	0.61	0.60	0.77	0.46
CCC	0.50	0.50	0.73	0.40
KDD	0.55	0.61	0.80	0.68

6.4 Structural model analysis

Smart PLS 3.2.7 statistical software was used to test the hypothesised model. PLS bootstrapping technique was adopted for resampling to enhance the prospect and the generalizability of the original sample data and analysis the structural model. As for establishing the measurement and the structural model fitness, three indicators were used as recommended by Hair et. al. (2014); beta, T-statistics and adjusted R2. The results of the PLS test are shown in figure 2, and 3.

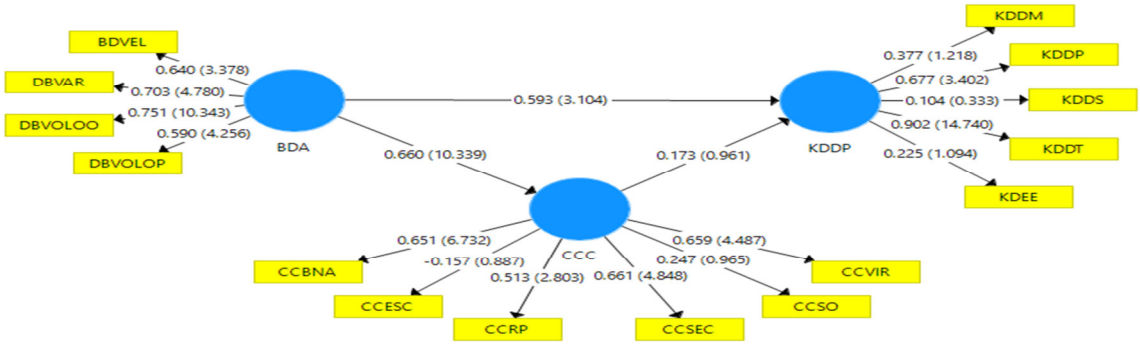

Figure 2: Hypothesised Bootstrapping Model-Direct Effect

For the Test Statistics (TTEST) of both inner and outer model to be accepted it should exceed 1.96 at the

0.05 significate level, all variable and regression paths that failed the TTest of significance were removed as in the modified model figure 4.

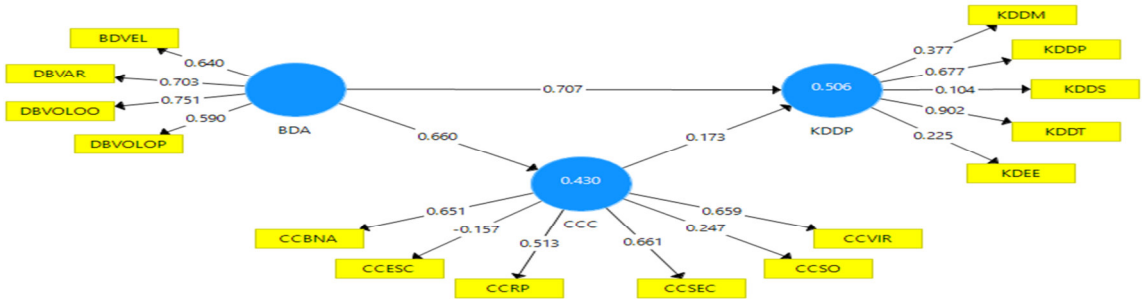

Figure 3: Hypothesised PLS Model-Total Effect

Table 7. Hypotheses Testing Results

Hypothesis	Beta	TEST	Adjusted R2	Decision
H1:There is a significant impact of big data attributes on knowledge discovery processes BDA ➡ KDDP	0.593*	3.10	0.50	Accepted
H2:There is a significant impact of big data attributes on cloud computing characteristics BDA ➡ CCC	0.660*	10.34	0.43	Accepted
H3:There is a significant impact of cloud computing characteristics on knowledge discovery processes BDA ➡ CCC	0.173	0.961	-	Rejected
H4:There is a significant mediating effect of cloud computing on knowledge discovery from big data BDA ➡ CCC ➡ KDDP	0.114	0.874	-	Rejected

* Significant on 0.05 significant level

According to overall testing results the decision was made to accept the H1, H2 hypotheses and to reject H3, and H4. Thus cloud computing characteristics has no mediation effect on knowledge discovery process from big data, and this results was also confirmed by Sobel test results showed in table 8.

Table 8. Sobel Test (Mediation Test)

Path TTest Value		Test	Test Statistics	P-Values	Decision
BDA ➡ CCC	10.1	Sobel	0.913	0.361	Rejected (insignificant)
CCC ➡ KDDP	0.92	Aroian	0.909	0.363	Rejected (insignificant

The researchers adopted Sobel test to further test and validate the mediation effect, as Sobel test is considered as a leading method of testing the significance of a mediation effect. According to Preacher & Leonardelli (2001), if p-value falls below 0.05 then mediation effect is significant, and as shown, p-values was greater than 0.05, hence there is no mediation effect of cloud computing between big data attributes and knowledge discovery process as suggest by the hypothesised model.

6.4.1 The modified model

Based on the results of the PLS test, in the three outer measurement models (BDA, CCC, and KDDP) and the inner structural model, only the variables that has t-test values (values in brackets) greater than 1.96 was accepted and kept in the suggested model. As to this, a number of variables were eliminated as they failed to have neither significant loading nor a reasonable explanation power to their dimension. A direct path (CCC - KDDP) was also eliminated as it fail to have any significate impacts. Figure 4 shows the remaining accepted variables in their dimension and two significate direct path of impact

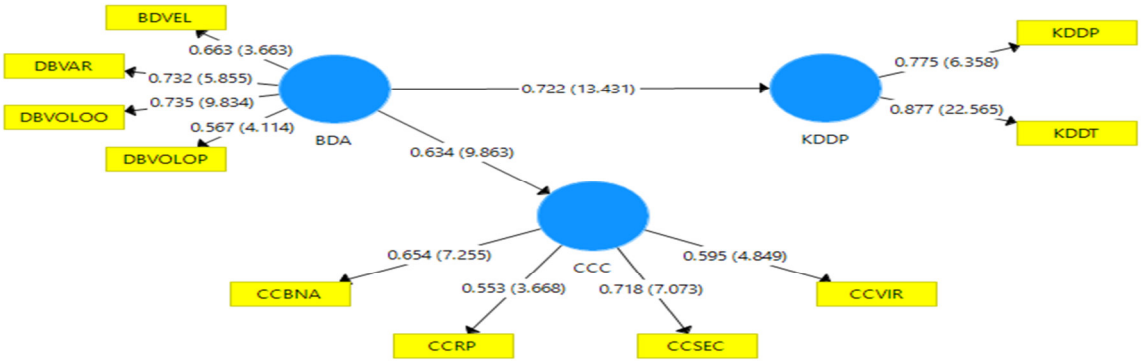

Figure 4: Modified Model

7.0 Discussion and Conclusion

This study attempted to investigate whether cloud computing can act as a facilitating and mediating agent in support of knowledge discovery from big data and the possible underlying effects and impacts between the proposed constructs. As for cloud computing characteristics, the results revels that only four out of six suggested characteristics can be accounted for and included in the possible structure of the cloud computing characteristics; broad network access, shared resource pool, security, and virtualization. As for big data attributes (3Vs), the results confirmed the attributes as suggested and introduced a miner amendment to the concept of volume were volume can be seen as a results of ownership of data and/or as a results of operations, where volume ownership can be seen as the main contributor to BD concepts. As for knowledge discovery process, the results only confirmed two factors; data process and data transformation, as the only ones that are effected by big data attributes. Big data attributes has a significate impact on data processing and transformation process of the knowledge discovery and in the same time big data attributes has a significate impact on cloud computing, as to this findings the study recommend that telecommunication industry and possibly the service industry should pay more attention to these characteristics and to investigate the activities that might add more value to and improve the robustness of such characteristics, attributes and processes as shown in the results of the study. These findings are expected draw more attention toward improving the outlook of such values as it can be employed to increase the yield and the production data and the discovery of knowledge.

References

Abaker, I., Hashem, T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., & Khan, S. U. (2014). The riseof “big data” on cloudcomputing:Review and open research issues. Elsevier, 98-115.

Alavi, M., & Leidner, D. E. (2001). Review: Knowledge Management and Knowledge Management Systems: Conceptual Foundations and Research Issues. JSTOR, 107-136.

Alkhaldi, FM & Qararah, MM. (2016) Cloud Computing Critical Factors and Investment Decision: An Empirical Investigation on Jordan Public Sector, Network and Complex Systems, Vol 6. No. 4., 5-10

Begoli, E., & Horey, J. (2012). Design Principles for Effective Knowledge. Tennessee: Oak Ridge National Laboratory.

Bezemer, C.-P., & Zaidman, A. (2010). Multi-tenant SaaS applications: maintenance dream or nightmare? IWPSE-EVOL '10 Proceedings of the Joint ERCIM Workshop on Software Evolution (EVOL) and International Workshop on Principles of Software Evolution (IWPSE) (pp. 88-92). Antwerp, Belgium: ACM New York, NY, USA ©2010.

Bhardwaj, S., Jain, L., & Jain, S. (2010). CLOUD COMPUTING: A STUDY OF INFRASTRUCTURE AS A SERVICE (IAAS). International Journal of Engineering and Information Technology, 60-63.

Durkee, D. (2010, April 1). Why Cloud Computing Will Never Be Free. Queue - Emulators, 8(4), 20. Retrieved 3 28, 2017, from http://dl.acm.org/citation.cfm?id=1772130

Feldman, R., & Dagan, I. (1995). Knowledge Discovery in Textual Databases (KDT). AAAI, 112-117.

Gullo, F. (2015). From Patterns in Data to Knowledge Discovery:What Data Mining Can Do. Elsevier, 18 – 22.

Jadeja, Y., & Modi, K. (2012). Cloud Computing - Concepts, Architecture and challenges. International Conference on Computing, Electronics and Electrical Technologies (pp. 877-880). Gujarat: 978-1-4673-0210-4112/$31.00 ©2012 IEEE.

Katal, A., Wazid, M., & Goudar, R. H. (2013). Big Data: Issues, Challenges, Tools and Good practices. IEEE, 404-409.

Lohr, S. (2012). The Age of Big Data. New York: The New York Times.

Mahmood, Z. (2011). Cloud Computing: Characteristics and Deployment Approaches. IEEE International Conference on Computer and Information Technology, 121-126.

Maimon, O., & Rokach, L. (2010). Introduction to Knowledge Discovery and Data Mining. Springer, 1-15.
McAfee, A., & Brynjolfsson, E. (2012). Big Data: The Management revolution. Harvard Business Review, 59-68.
Mell, P., & Grance, T. (2011). The NIST Definition of Cloud computing. Gaithersburg: National Institute of Standards and Technology.
Qusay, H. (2011). "Demystifying Cloud Computing" (PDF). The Journal of Defense Software Engineering (CrossTalk) 2011 (Jan/Feb).
Ristoski, P., & Paulheim, H. (2015). Semantic Web in data mining and knowledge discovery: A comprehensive survey. Elsevier, 1-22.
Hair, J., Anderson, R., Tatham, R., and Black, W.; (2010). Multivariate Data Analysis (7th edition). Prentice Hall International, Inc Upper Saddle reviver: New Jersey - USA.
Hair, J.;, Tomas, G., Hult M., Ringle,C. and Sa rstedt M., (2014) A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM), SAGE Publications, Inc.
Hurwitz, J.; Bloor, R.; Kaufman, M., and Halper, F. (2010). Cloud Computing for Dummies. Wiley. ISBN: 978-0-470-48470-8
Takabi, H., & Joshi, J. B. (2010). Security and Privacy Challenges in Cloud Computing Environments. THE IEEE COMPUTER AND RELIABILITY SOCIETIES, 24-31.
Wilson, T. D. (2002). The nonsense of 'knowledge management'. Information Research, 144.
Zhong, N., & Skowron, A. (2001). A ROUGH SET-BASED KNOWLEDGE. Journal of Computer Science, 603-619.