ENHANCED CLOUD DATA MANAGEMENT MODEL FOR THE INTERNET OF THINGS

أ. اماني مصباح فرفر Amani M. Farfar

أ. حلمي المختار سالم Helmy M. Salem

أ. عبدالمنعم علي خليفة Abd ALmonim A. Kalifa

المعهد العالي للعلوم والتقنية- الاصابعة

amanifufur@gmail.com / 0913533494

الملخص

أدت الاتجاهات الناشئة لإنترنت الأشياء في عالم اليوم إلى إعادة اختراع أفضل التقنيات الممكنة المتضمنة في أنظمة إدارة البيانات السحابية. وبالمثل، لا يمكن إهمال احتياجات تقنيات إدارة البيانات المتقدمة وأهمية نماذج إدارة البيانات السحابية المحسنة. يعمل العملاء والمستخدمون النهائيون والمهنيون في الشركات المضيفة بجد لتلبية متطلباتهم اليومية المتعلقة بالبيانات، ويحاولون التعامل مع الكميات الهائلة من البيانات التي يتم إنشاؤها نتيجة للأجهزة والتطبيقات الذكية لإنترنت الأشياء، وعادةً ما تعتمد التقنيات على الخوارزميات وعمليات إدارة البيانات المختلفة ضمنيًا بدلاً من الأجهزة باهظة الثمن.

ومن ثم، تم اقتراح نموذج معزز لإدارة البيانات السحابية لإنترنت الأشياء بعد دراسات مستفيضة للمقالات البحثية والتحليل الدقيق للبنية التحتية الحالية والثغرات التي يواجهها قطاع تكنولوجيا المعلومات، وقد تم تحقيق ذلك من خلال المراجعة والاستبيان المصمم للحصول على الإجابات ذات الصلة حول المشكلات والاقتراحات. يمكن أن يكون النظام المقترح قابلاً للتكيف بسهولة حيث يمكن استخدامه بشكل فعال في البنى التحتية الحالية لأنظمة إدارة البيانات المنشورة بالفعل.

وقد تم اقتراح النظام بناء على النتائج التي توصل إليها الباحثون ومقترحات الخبراء الفنيين التي وردت نتيجة لردود الاستبيانات من الخبراء الفنيين. ومن ثم سيسهم في حل مشاكل إدارة البيانات المتعلقة بأجهزة وتطبيقات إنترنت الأشياء.

الكلمات المفتاحية: تكنولوجيا المعلومات (IT)، إدارة البيانات السحابية (CMD)، إنترنت الأشياء (IoT)، تحديد الترددات الراديوية (RFID)، شبكة الاستشعار اللاسلكية (WSN)، معالجة الأحداث المعقدة (CEP)، البنية التحتية كخدمة (IaaS)، البرمجيات كخدمة (SaaS)، البيانات كخدمة (DaaS)، النظام الأساسي كخدمة (PaaS).

Abstract

The emerging trends of Internet of Things in today's world have given rise to reinvent the best possible techniques implied in Cloud Data Management systems. Similarly, the needs of advanced data management techniques and the importance of enhanced cloud data management models cannot be neglected. The clients/end-users and professionals at hosting companies are working hard to fulfil their day-to-day data related requirements, and trying to handle the huge bulks of data being generated as result of IoT smart devices and applications. Normally the techniques are based on the algorithms and different data management processes implied rather than the expensive hardware involved.

Hence, an enhanced Cloud Data Management model for Internet of Things has been proposed after thorough studies of research articles, and careful analysis of the current infrastructure and loopholes being faced in IT sector. This has been accomplished by the literature review and the questionnaire designed to obtain relevant responses about the problems and suggestions. The proposed system can be easily adaptable as it can be actively used on existing infrastructures of already deployed data management systems.

The system has been proposed according to the researchers' findings and suggesti ons of the technical experts received as a result of the questionnaires' responses from the technical experts. Hence, it is believed that it is very much likely to solve the data management problems related with IoT devices and applications.

Key words: Information Technology (IT), Cloud Data Management (CMD), Internet of Things (IoT), Radio Frequency Identification (RFID), Wireless Sensor Network (WSN), Complex Event Processing (CEP), Infrastructure as a service (IaaS), Software as a Service (SaaS), Data as a service (DaaS), Platform as a service (PaaS).

Introduction

The Internet of Things has undergone a rapid development which combines integrated sensors, communication links, devices and other analytic technologies. The revolution has not only contributed to the abandonment of the traditional methods of data processing but rather a greater improvement in generating, collecting, managing and analyzing of information. Some of the emerging technologies that exist in the internet of things do include: - Radio Frequency Identification (RFID), Wireless Sensor Network (WSN) and Complex Event Processing (CEP) making it easier to monitor, react accurately to the physical world. Although the data captured in the Internet of Things is growing in numbers unexpectedly, Angarita (2015) asserts that companies and organizations should embark on new techniques and ways that include improving infrastructure thus enhancing data management capabilities and also addressing various challenges and opportunities in the Internet of Things. The Internet of things is one of the trendy computing domains which allow devices, machines and sensors to autonomous communicate amongst each other. Some of the existing free cloud storage applications do include: - Drop box, Flip drive, Google Drive, One Drive, P Drive among others. The cloud is integrated with so many services as identified in the cloud computing model for example: Infrastructure as a service (IaaS), Software as a Service (SaaS), Data as a service (DaaS), and Platform as a service (PaaS). Among these services, organizations and other business enterprises need to utilize the Data as a service (DaaS) in order to make better decisions from the Data. The data currently accessed is from multiple sources called the internet of things which are independent and distributed all over the world for example: Web services like Facebook, Twitter, YouTube and Flickr among others. Gartner reports highlight that the cloud computing deployment model is a driver in supporting the IoT Technology. Some of the models do include: Private Cloud, Public Cloud, Community Cloud and Hybrid cloud. Since the Internet of Things is associated with the cloud itself it is very important to address the models where data can be essentially stored.

Problem Statement

Like never before, data management has become a critical issue in this recent era and yet received less attention. Firms and enterprises have been compelled with massive data that needs to be harness in order to gain significant insights. Although various data strategies have been proposed to leverage retrieval and access to data, the various data management models are not capable enough to be used in the Internet of Things Gang (2010). Similarly Krensky (2015) identifies that not every enterprise is ready to invest in data storage housing facilities to cater for the IoT data thus carryout prioritization of what kind of data can be collected from the network. This would mean that data has to only be prioritized to be backed up in the current architecture of the data centers yet the heterogeneous nature of data is so overwhelming for enterprises.

Aim and Objectives of the Research

The aim of this research is to investigate the concept of cloud data management in the Internet of Things (IoT). Secondly to propose an enhanced data management model for the Internet of Things. The objectives of the research are therefore as follows:

1. To study the concept of cloud data management in the IoT and its associated challenges.
2. To investigate the various cloud data management models for the Internet of Things.
3. To propose a cloud data management model that suits the Internet of Things.
4. To Test and Evaluate the proposed model.

Significance of the Research

The research study will contribute to the academia, researchers and various enterprises in the following ways:

1. Various researchers have been able to propose various data management approaches and models however, there is still insignificant information on how to leverage and harness the massive IoT Data.
2. Enterprises have not been able to come up with appropriate mechanism to store all the data that comes in from the various IoT networks, there as a result the research will provide a flexible and better approach to deal such constraints.
3. Conducting this research will also broaden on the knowledge and published content regarding data management in the Internet of Things since less works has been done in this area.

Scope of the Study

The research study mainly focuses on data management in the internet of things, the research will technically focus on data management approaches applied in the cloud visa-vise the Internet of Things. It is therefore imperative that the research concentrates on the Internet of Things paradigm as it raises more concerns with regards to data management strategies and approaches.

Research Questions

The researchers were compelled to develop a few research questions to address the research problem. These research questions act as a bench mark to identify whether the aims and objectives have been fulfilled.

1. To what extent can IoT data be appropriately managed in the cloud?
2. What data management strategies should be applied in order to harness IoT data?
3. How can we improve on the data management capabilities in the IoT network?

Literature Review

As technology continues to evolve, various technological paradigms have emerged with more sophisticated computing power. Currently, the internet serves as fundamental bedrock and also a driving force towards emergence of new technological paradigms. The most recent technological paradigm with high computing prospect are cloud computing and also fog computing (Bonomi et al., 2012; Saharan, 2015; Ruoss, 2016). These two technologies carryout high computing process with massive data storage capability and also ensures easy accessibility of such data. Cloud computing and also for computing has ensured that the technological application and processes of internet of things is ensured appropriately. They currently serve as a major driving of technological processes in our environment and also equip and support operational capability of IoT (Bonomi et al., 2012; Belli, 2015).

Related Work and Research

Cloud computing and IoT technology are often described as ubiquitous computing. Cloud computing is described as a technological concept which mitigates storage challenges faced in many sectors of our environments, it eliminates the need to own huge in-house data storage facilities which can cost a lot to maintain and also secure. When cloud computing is combined with Internet of Things, it leads to a connection between data and information. Hence cloud computing is said to be possess the computing strategy that can meet the needs of data management for this era (Rui & Danpeng, 2015). Cloud computing paradigm involves the storage, analysis, processing and retrieval of huge heterogeneous data for easy decision making process for an organization or individual (Kshetri, 2012; Alzahrani et al., 2014; Smith et al., 2014). Internet of Things is verily an emerging field in modern era and plays a vital role in the interconnected world of computing. The IT statististics and inclinations reveal that internet now plays a major role in everyday life and things (Thomas, 2016). The internet of things (IoT) which is described by (Lee & Lee, 2015) as internet of everything and is recognized as a very strong driving force in virtually all sectors of our environment. The rapid increase in the growth of WiFi and 4G is evidence of the emergence of ubiquitous computing and information sharing over the networks (Gubbi et al., 2013).

Research Methodology

Research is defined as discussing and exploring a specific problem or topic by use of a systematic, strategic or scientific approach. It basically refers to the search, re-search and explore the issues and investigate deeply to study the possible outcomes. Petersen & Gencel (2013) assert that both the Quantitative and Qualitative approaches to data collections have been widely accepted by most researchers. It is also clear that both the approaches can be used together thus referred to as mixed methods. The quantitative approach involves sampling large populations and usually through use of questionnaires.

Research Approach and Design

Gubbi et al. (2013) state that research design is the fundamental building block on any research. It assists in giving a structure to the entire research and categorizes the methods in which various parts of the research work join together to answer the research question(s). In this research work, exploratory study has been conducted beforehand, in order to critical analyze the research problems and get clarification about the research objectives. Exploratory study includes investigating the area of the problem at hand and provides the researchers a detailed inner view of the situation by the generation of basic knowledge and understanding using methods e.g. questionnaires and in-depth group or individual studies and discussions with experts for attaining the best possible clarity (Krensky, 2015).

Second form of research design used is descriptive research which comes into existence only after in-depth exploration and organization of the findings and checking or verifying the explanations aligning with the main causes for carrying out research which are for description, explanation and validation. For this research work, researchers started with brain storming sessions and going through the research articles and blogs available for cloud data management, and after analyzing its characteristics and benefits, Internet of Things paradigm has been explored, and issues and challenges faced by CDM for IoT were studied and resulted in the clarification of research problem and objectives.

Research Approach

There are different approaches to research which include the procedures and techniques from comprehensive concepts to minute details of the data collection methods, data examination and analysis and data interpretation. The research approach strategy comprises of multiple conclusions, which can be achieved in any logical order which is as per the demonstration of research. However, it is a vital verdict to choose the approach of exploring a research subject Cisco (2016). For this research work, the researchers explored multiple ways to carry to carry research and chose he mixed approach as best fit to carry out with data collection and analysis using questionnaire survey and literature review. The chosen research philosophy can be one amongst Pragmatism, Positivism, Realism or Interpretiism (Fan 2010). After comparing different philosophies, the most suitable philosophy for this research found is pragmatism which follows mainly the quantitative approach. For this research, the researchers chose the Pragmatism approach, as it involves both the qualitative and quantitative research methods, which are the requirements for this research work. The primary data has been collected through questionnaire which includes both open-ended and close-ended questions. Below are described the quantitative, qualitative and mixed approach methods applied for this research.

Collected Data Analysis

Data has been collected and analyzed through both quantitative and qualitative approach using graphs, analysis tables and charts for basic information and affinity analysis for open ended questions.

For the detailed data analysis, the responses have been appropriately understood and interpreted. Initially the collected data is checked for shortcomings and filtration is done so that a detailed analysis can be carried through. This assisted in getting to know the issues present with the responses of the questionnaire, for example, many questions were left unanswered by the respondents. To solve this, the researchers reverted back to the individual to get proper response and in case of failure the particular response has been discarded for analysis. Also, the filtered data is further checked to sort out for questionnaires that may have a lot of incomplete responses. Afterwards, results have been discussed for each question one by one by categorization of the data in the form of chart, leading to the grouping of a like findings into logical chunks and resultant information has been highlighted for further analysis as explained in affinity analysis chart.

Analysis of Respondents' Characteristics

Questionnaires have been distributed via Google Forms to a number of technology experts in Kuala Lumpur, Malaysia's main Information Technology (IT) center. Some of the questionnaires were handed to the respondents by hand, which mainly included CEO's/ CTO's, Project Managers/ Coordinators, Consultants, Data Scientists, Business Analysts and other such persons who work directly and daily with cloud management systems and deal with storage capacity issues on their job routines in software engineering/development (small and medium level) companies. The respondents were chosen because they have been serving in various private companies in IT sector to provide best software and technical solutions to their clients for years, and have to deal with different client requirements on daily basis.

The questionnaire has also been shared with multiple regions across the globe, according to our acquaintances. Different software and technology experts have provided the feedback according to their knowledge, experience and on-the-job requirements. In the questionnaire, multiple questions have been asked to obtain an idea about their expectations from the proposed solutions from this research. The varied feedback has been helpful in realizing the lookout of the users and their needs, also the flaws and discrepancies faced by the currently deployed systems. The experts have also given a few suggestions for better enhancements of CDM for IoT.

Respondent Specifications

Role in Organization

The following table presents the percentage of respondents according to their position in the organization. It is depicted that the respondents to research questionnaire were 9% managers, 12% coordinators, 13% consultants, 37% developers, 12% chief technology officers, 7% data scientist and 10% business analysts.

Role in Organization	Respondents Percentage
CEO/ CTO	9%
Project Manager/ Coordinator	12%
Consultant	13%
Data Scientist	37%
Business Analyst/	12%
Other	17%

Figure 1 Aggregation based on Role in Organization

Gender

Our group of respondents included both males and females but majority of them were males. 74.7% of the respondents were male and 25.3% were females as shown in figure below. Assumption has been made that usually females are working in Human Resources and Design related jobs and males are more in the sector of I.T. related jobs.

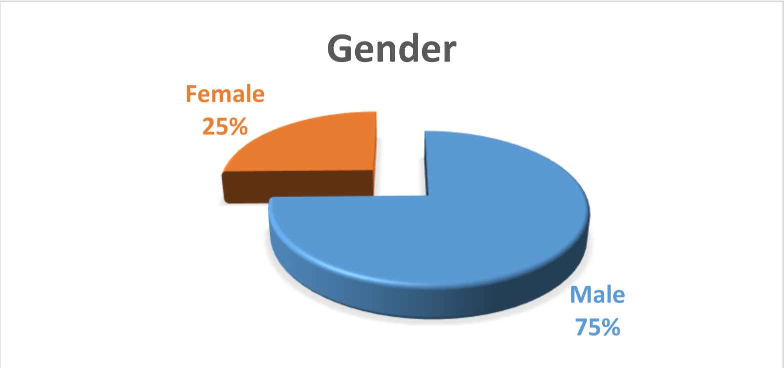

Figure 2 Gender-based Aggregation of Respondents

Experience Level

According to the responses, most of the respondents had 5-10 years of industrial experience. Only 15.6% of the respondents had less than 5 years of experience, while 77.5% respondents had experience varying from 5 to 10 years. Senior respondents having experience between 11-20 years were 5.5% of the total while 1.4% of them had experience more than 20 years. This shows that majority of those using Cloud Data Technologies are sufficiently experienced professionals. This has assisted in obtaining more precise responses as the respondent group has a vast practical and daily life experience of dealing with Cloud Data Services and Internet of Things.

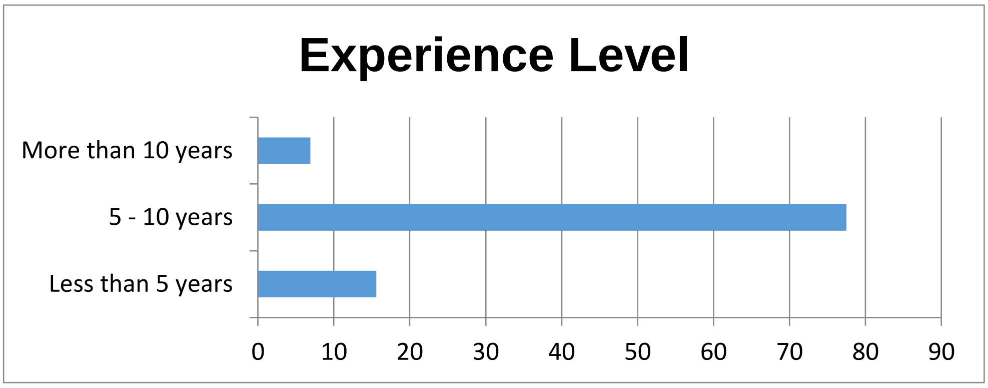

Figure 3 Stacked bar depicting the experience level of respondents

Qualification

All of the respondents are educated and highly professionals who had good academic background as the statistical analysis shows 47.5 percent of the respondents have at least Graduate and 52.5 percent of the respondents have already completed Master's degree or above and have vast experience in the respective field. The section A contains general questions regarding experience level, gender, qualification, role in organization and the affinity analysis, as it shows that most of the survey respondents have good understanding and experience of Cloud Computing and Internet of Things.

Discussion

The questionnaire has been given to respondents in Kuala Lumpur, Malaysia's main Information Technology (IT) center. The respondents include professionals working in software engineering/development small and medium level companies. The questionnaire has also been shared with multiple regions across the globe, according to our acquaintances. Different software and technology experts have provided the feedback according to their knowledge, experience and on-the-job requirements. In the questionnaire, multiple questions have been asked to obtain an idea about their expectations from the proposed solutions from this research. The varied feedback has been helpful in realizing the lookout of the users and their needs, also the flaws and discrepancies faced by the currently deployed systems. The questionnaire was mainly distributed online via email and google forms; while few questionnaires were distributed among professionals in some IT companies in Technology Park Malaysia in form of hardcopies by hand. The researchers also found some incomplete responses from respondents which has not been included in the analysis. The affinity analysis has been used to analyze the descriptive responses.

Affinity analysis

Affinity analysis is an analytic technique that helps organizing and managing data relatively in form of groups or individuals. This technique has been found useful when there is lot of random data from studies, surveys, close or open-ended questionnaires and interviews, in order to produce a meaningful analysis of data (Thomas et al. 2017).

Affinity Chart

As described by Zang et al. (2010), Affinity diagram has been used to organize information compiled during analysis session of descriptive responses, and to organize ideas or attributes. It has basically combined the co-occurrences of multiple activities or responses from different individuals in response to a certain question. An Affinity diagram is designed below for the analysis of the responses obtained from Section B of the questionnaires.

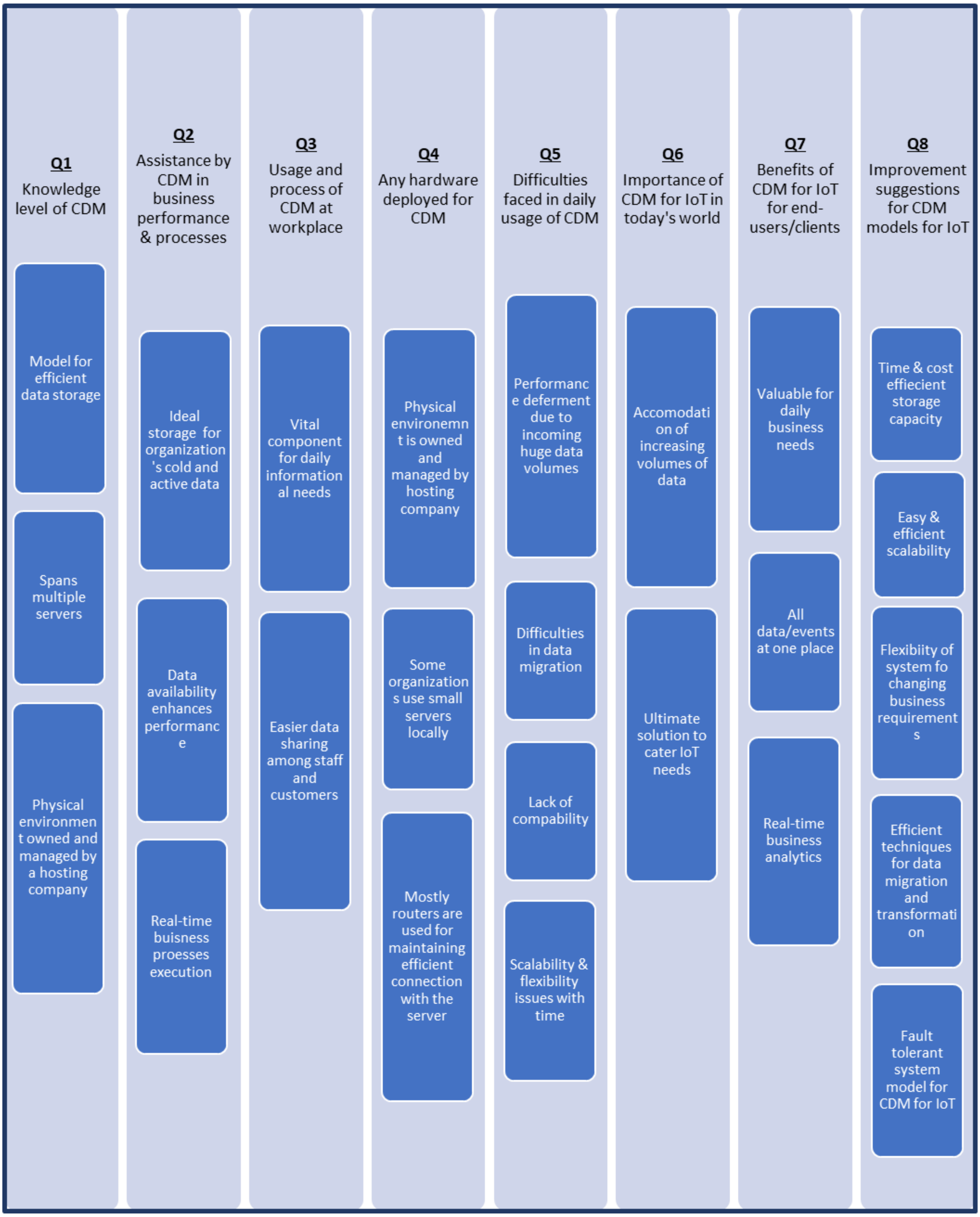

Figure 4 Affinity Analysis Diagram of Questionnaire Responses from Technical Representatives (Self-designed)

A wide range of capabilities are being offered by storage systems in terms of capacity, data protection and cost. Q1 reveals that people have enough knowledge about

Cloud Data Management, and understand the importance of how things work in cloud data management to much extent. But there are technical officials, who lack the understanding of the real usage, technicalities and complexities of CDM, and hence needs some orientation and training sessions in this regard for better understanding and better data management.

Data protection on cloud systems are being employed by almost all companies on small, medium or larger scales. Data management has become the need of the day, hence from Q2 responses, it has been inferred that CDM provides an ideal storage for the organizations' cold and active data storage, hence enhancing the processes and performance by readily data availability in this way.

It is evident from Q3 that Cloud Data Management serves as an important component for the business daily informational needs, provides the data standards met and the flexibility ensured in terms of protection, capacity, tuning and maintenance time and cost of such a system. Most of the organizations are availing services of big cloud hosting third party companies. They host all their data on cloud in order to achieve much more stability and reliability for all their data related needs. Generally, the physical servers and advanced hardware are placed with the hosting companies while the clients' devices may or may not be very advanced as they just have to maintain an efficient connection with main server.

Q5: serves as a very important question for our research topic. The technologies are very advanced and efficient in the beginning, and companies purchase an ideal storage to fulfil their data related needs. But such ideal cloud storage techniques could not stand the incoming data volume and hence result in the performance deferment. Initial difficulties include transfer of data from older servers to the new cloud servers, as data integrity and stability as well as events streaming and storing has to be maintain at every level. The data once transferred to a new cloud are usually not compatible and flexible for the ever-changing requirements of the company. And the data administrators keep on working hard to let the applications using data from clouds not to suffer performance and time issues. The respondents see the challenges of CDM usage in perspectives of scalability, integrity, time and cost effectiveness.

According to the responses we obtain from Q6, it has been inferred that the respondents are very much curious about the upcoming technologies to cater the increasing volumes of data in today's world. Most of the respondents have analyzed the importance of Cloud Data Management as an ultimate solution to cater the needs of Internet of Things in terms of storing all the data and events, run the important queries to extract data and to run analytics to get insight into stored data. The clients/ end-users visualize CDM as valuable to their business needs. In response to Q7 from questionnaire, respondents agree the benefits of CDM for analysis of real-time data and hence detect and react to the real-time business opportunities and threats to the business. The scalability, flexibility and fault tolerating capability of Cloud Data Management for Internet of Things helps in carrying out routine business tasks, practices and processes smoothly and without delay.

When asked about the suggestions for the improvement of existing Cloud Data Management models for IoT, respondents gave valuable suggestions in Q8. Most of the technical representatives declared that they need more time and cost efficient storage capacity, which should be easily scalable and flexible to meet changing business requirements. Other respondents recommended that there should be introduction of more efficient techniques for data migration and transformation. On-job technical experts also mentioned the need for a fault-tolerant system model for CDM for IoT mechanism.

It can be concluded that with the prevailing benefits of Cloud Data Management for Internet of Things, there are arising shortcomings that need to be cater for running routine life slickly in the coming days of increasing bulks of Big data. Hence, researchers and Cloud hosting companies need to work on the enhancement of their servers to cater the needs and demands of clients with provided flexibility, scalability and cost effective solutions to the data challenges in the world of the Internet and the interconnected devices and networks.

The results of the questionnaires as a whole depict that the adoption of a new and enhanced Cloud Data Management Model, particularly for Internet of Things to increase the efficiency and smooth run of business processes. The enhanced model should work for the improvement in reducing the time taking and resources consuming processes carried out by several companies. The successful adoption of the enhanced CDM model for IoT will allow easy data management in less time, automated resources and processes and increased production. It is expected to increase the quality of service being delivered as well as being relief of technical staff from repeating hectic tasks on daily basis. It will make easier the data transformation, integrity, queries processing and cost effective storage scalability. Hence, the data management can be automated to a greater extent by using the proposed prototype as a base. Additionally, the results of this research comply with the previous studies and researches and surveys done in the big data and cloud data management fields to meet the requirements of the clients and users with the Information Technology sector and effectiveness of the cloud storage systems to reduce delays in consumer services.

Proposed Model and Evaluation

The level of data generated on a daily basis continues to expand due to advances in different technological paradigms; the data generated forms important fundamental aspect of many sectors in our environment. Due to sensitivity of most data, many research works have proposed that measures, techniques, models and framework be developed to ensure that data needed for decision making process are adequately managed and processed. According to (Ji et al., 2012). Researchers have been putting effort to minimize the tailback in traditional database management systems and to replace them with the new technology of cloud data management in order to continue serve and maintain the incoming bulks of data.

As discussed in the literature review about the capability of IoT systems to produce enormous amounts of data, it is imperatively important that a data management system or model is proposed or developed to handle the amount of data generated by IoT systems and devices. The proposed model or system should be able to address the limitations faced by conventional data management used by IoT systems, based on the literature review, data confidentiality, data integrity, availability, real-time processing, data accuracy, data migration, data cleaning and fault tolerant management systems are some of the issues identified. To ensure that IoT systems properly functions with sound management, it is pertinent that the identified limitations are properly addressed.

The responses of the questionnaires provided from technical experts combined with the literature on cloud data management systems for Internet of Things have been used for purpose of proposing a prototype that can assist in development of an improved CDM model for IoT.

Literature Review for Design of the Artifact

A comprehensive literature review has been written covering maximum aspects of the past related work carried out globally in designing data management models and their implementation with different priorities. The major focus of this research is on providing assistance to design and develop more efficient cloud data management model for IoT. The reason for selecting this topic is because the issue of receiving more and more bulks of data day by day and the difficulties of efficient managing of data from multiple sources has become a challenge for IT experts.

Most of the literature has placed emphasis on the importance and limitations of the existing data management models and how internet of things is playing a major role in the excess generation of data from the globe over the Internet. Furthermore, analysis done on the existing data management models and systems has also provided clues for easing the process of proposing this prototype. The proposed model cloud data management model for IoT is proposed based on tailoring of (Abu-elkheir et al., 2013) IoT data management model. Although the inspiration has been taken from a similar system already implemented, yet some differences are there as the newly proposed prototype is better in various ways. The said study focuses on how the data is collected, stored and processed, whereas this research works focuses on the scalability, fault tolerance and availability issues mainly and data processing and storage as secondary to come up with the best solution for the primary research points.

Assistance of Affinity Analysis in designing the artifact

The responses obtained from the questionnaires have assisted a lot in coming up with an effective artifact. Some important factors have come in front for which attempts will be made to include in the prototype.

The currently used environmental practices unleashed in the analysis are perfect for implementation of an enhanced cloud data management system for IoT, hence the researchers are confident that the proposed system will be able to blend in and prove very useable on the devices currently used in the observed organizations. These devices include laptops, tablets, desktop pc’s, routers, servers, drives, and bandwidth etc. All the respondents suggested the improvements in CDM for future use and compatibility and these suggestions have been taken into account for proposing a best possible model, considering the need of the day in this aspect.

Proposed Cloud Data Management Model for IoT

A framework has been proposed in this research for enhanced cloud data management model for IoT assisting technical experts to use the available latest technology in an efficient way. This will most probably bring relief for the hosting companies, while bringing more business for them. The system proposed will be useful for both the hosting companies and the end-client users in several aspects. Main reason behind proposing this framework is to assist efficient data management on the cloud. The enhanced model has been depicted in the figures below for the purpose of the presentation of system including general overview of cloud data management, architecture of the proposed prototype and the lifecycle of the proposed system.

General Overview of CDM

The figure below depicts a general overview of Cloud Data Management and how things come in and go out from cloud servers. Basically, there are three different components. It has been shown in figure that there are several data sources from where data is supposed to be collected and aggregated in cloud. The data is then processed, cleaned, managed and stored at the cloud servers. The data quality is maintained and data encryption is implemented to obtain data security primitives. The data acquisition and virtualization is done at the third component to provide query optimization and generate analysis graphs and charts according to the requirements, also the correct, reliable, timely and accurate data availability is maintained at this end. All these CDM components ae linked with each other to keep things working and linked.

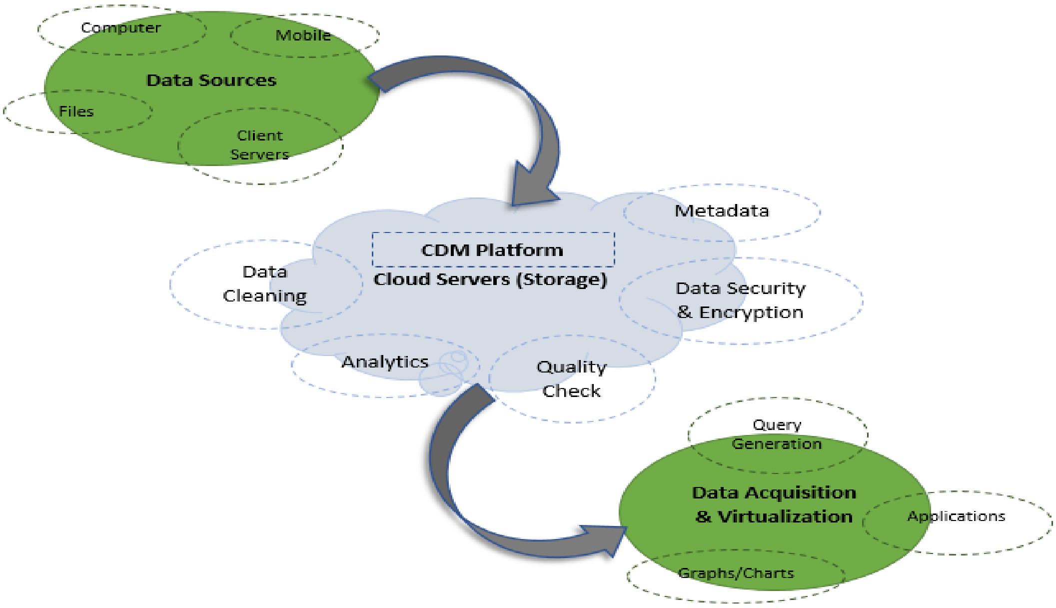

Figure 5 General Overview of CDM

Detailed Architecture of Proposed Model

Detailed diagram of proposed architecture of enhanced cloud data management model working with IoT devices has been shown in figure below.

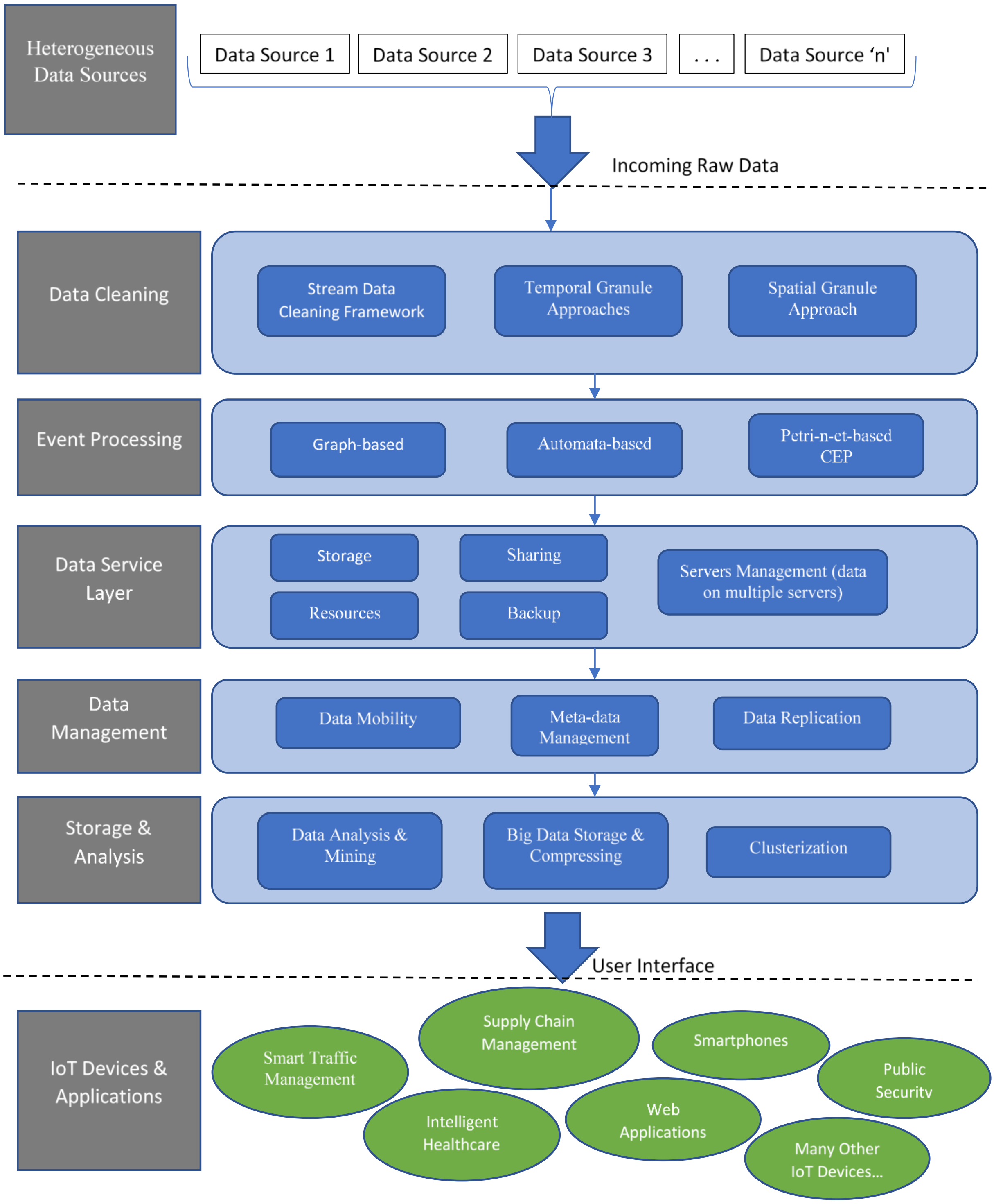

Figure 6 Architecture of Enhanced Cloud Data Management Model for Internet of Things

The process begins when data is generated from several multiple heterogeneous sources and this raw data then has to be moved to the cloud for efficient storage and processing. Below are explained the working of each layer of the proposed model as shown in the figure.

- **Data Cleaning**

This layer is necessitated due to the ability of IoT systems to generate huge amount of data within time intervals. Given these huge amounts of data generated by IoT systems with heterogeneous sensors and smart objects, there is bound to be significant amount of error, these errors pose great challenges in IoT functionality. Due to the errors generated by IoT systems and devices, the sole purpose of this layer is to mitigate all data unreliability that can result in during data generation and transmission. Also, this layer is characterized as the data cleaning because it serves as a platform for proper interpretation and analysis of IoT data. The temporal granule and spatial granule approaches have been used for cleaning and reprocessing of data.

- **Temporal Granule Approach**

This technique for data cleaning or reprocessing is based fundamentally on the idea of time window smoothing, it ensures that data is clean based on statistical characteristics. Temporal granule as the name implies is considered as a low-level clean up technique in which to achieve higher data cleaning or reprocessing, an additional spatial and also logic data is required.

- **Spatial Granule**

The main essence of a spatial granule in the data cleaning process is to simply clean, process and aggregate the data from various IoT sources. Considering that the technological concept of IoT revolves around readability in heterogeneous smart devices, the spatial granule technique or process ensures that the various data collected from various IoT devices and smart objects are properly refined and cleaned to ensure smooth execution of commands.

- **Event Processing**

The event processing layer is considered to be highly reliable and also has good usability, it is a layer characterized to transform and inferences unprocessed data into higher-level logic for organizational purposes. This layer processes raw data because raw data only gives simple information, often not related to business processes of the organization. The layer includes the graph, automata and petri net-based processing of data. Graph-based processing is a technique for parallel streaming of data and is efficient to process bulks of data streams in a timely manner by splitting them into independent parts. Automata based event processing approach has been applied for dynamic data and to determine the complexity of data with reference to state transitions, implementing the dynamism and adaptness on the incoming data. It provides complex reasoning about data processing by processing complicated streaming of data. Petri net-based model has been used for service composition, as it uses the 'rtc' function for performance and takes in account the reliability, response time and cost.

- **Data Services Layer**

Data services layer initiates the process of storage by allocating resources and implement the sharing of these resources between different types of heterogeneous data. This step provides the backup of data stored at different servers. Servers are managed and to enhance data storage and fault tolerance capability of cloud, data is placed at multiple low-scale servers rather than utilization of one big and expensive server which is unaffordable sometimes.

- **Data Management**

At the data management module data is managed in such a way to make data mobility easier, that is, to make the data compatible enough for the laid-back movement of data from one cloud to another. Meta-data is managed and indexing is done for easy, speedy and reliable access of data on the cloud. For the data security enhancement, several data nodes are created and data is replicated to quickly recover the data in case of failure of any component via help of advanced meta-data management and replication.

- **Storage & Analysis**

This layer ensures that all useful data generated are properly stored for future accessibility of the data. The services of IoT ensure an appropriate schema or data storage, to accommodate maximum data with the usage of minimum space. This will also lead to an improved analysis with various multiple attributes and hierarchies.

Clusters of data formed based on several similarities and dissimilarities based on types industries, applications and devices.

The last but not the least comes the IoT devices and applications like smart transport and traffic management systems, healthcare systems, smartphones, web applications, supply chain management systems, public security applications and many other such devices and applications. These IoT devise and applications access data through a user interface and via communication channel composed of networks like GSM, 3G, 4G, and network operators provided wireless networks and internet connections.

5.4.3. Proposed Data Management Lifecycle

The proposed data management lifecycle model is formulated to depict the major steps and process of how the enhanced cloud data management model works. The lifecycle is shown by figure below.

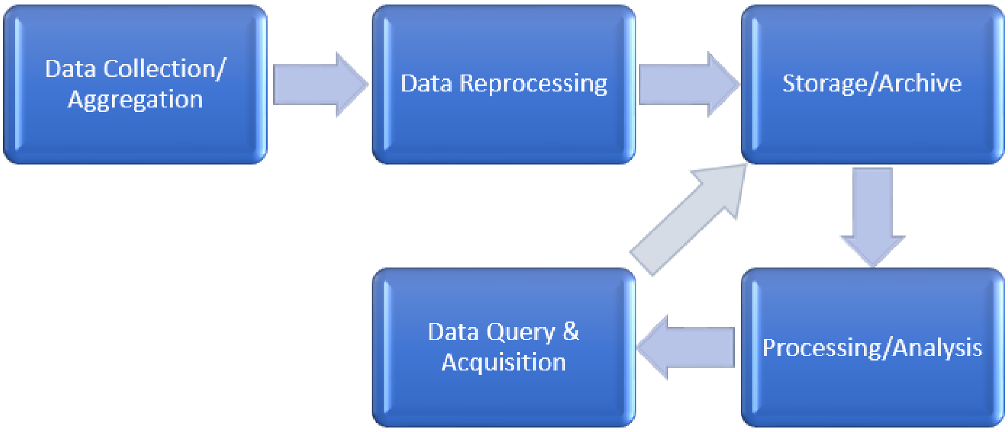

Figure 7 Proposed Data Management Lifecycle Model

- **Data Collection/Aggregation**

At this module, it is pertinently important that sensors as well as smart objects data are collected at time intervals. This data can be collected at various network points or gateways for filtration and analysis. Since the IoT is characterized as a technological paradigm which constitutes wireless communication technologies, technological computing devices such as WiFi, RFID, Zigbee, and cellular infrastructures constitute major components used in data transmission that is sending, receiving and collecting of data. These technologies are known to transmit in real-time and due to bandwidth limitations and expensive data streaming, raw data transmission over the network is typically avoided. Due to this fact from the aforementioned researchers, it is pertinent that techniques or methods are deployed for summarization and also for merging operations in real-time to increase efficiency in data management of IoT systems and also compress the amount of data collected for storage.

- **Data Reprocessing/ Cleaning**

This module is paramount in the management of IoT data because the technological paradigm of IoT involves enormous collection of data from heterogeneous devices, systems or things with varying format as well as structures. Due to this fact, the need for data reprocessing becomes paramount, data collected must be reprocessed so as to handle all missing data, eliminate redundancies and also properly integrate the data from different IoT sources into a more dynamic and unified schema. At this phase, margin for data uncertainties have been catered at this phase in order to avoid a fatal error that may be caused later due to any discrepancies left in raw form from any of the data sources.

This process is regarded to be paramount because it also constitutes a major procedure ensured in data mining which can be further described as data cleaning. Schema integration refers to a dynamic and abstract definition of a profound way towards accessing data without the need to customize data access for every source of data formats. The techniques used for data reprocessing include are temporal granule approach and spatial granule approach, which have already been discussed in detail in previous section.

- **Data Storage/ Archive**

This layer of the IoT data management life-cycle model ensures the effective and efficient storage and organization of data through a continuous updating the stored data with new data at intervals. The process of archiving is considered important in the proposed life-cycle model due to the fact that proper command and data storage is ensured. Archiving is described as the process of storing data offline for long-term storage which might not be urgently needed for the IoT systems ongoing operational functionality. The fundamental core in centralized data storage involves the development and deployment of a defined storage infrastructure that must be able to adapt to the heterogeneous IoT data types with real-time frequency. In a much later stage, relational database management process will be involved in IoT, it involves the management and organization of data into a tabular schema through a predefined interrelationship with metadata to ensure effective retrieval of data.

- **Data Processing/ Analysis**

This module is characterized with constant retrieval as well as analysis operations ensured on stored and archived data in a bid to ensure proper insight into historical data, to also predict future trends, and ensure no abnormality in stored data. Data processing is necessary due to the need to properly interpret, manipulate and transmit useful data for IoT, this is to ensure proper service delivery in operational functions, it also entails that all data supplied to various components of an IoT systems are correct and also relevant to the operations. This process involves appropriate techniques or technologies to ensure proper classification of data commands into hierarchical sequence of importance.

- **Data Query**

IoT rely enormously on data querying as a core process towards the accessing, retrieval and processing of data. An IoT query may ask to retrieve access to real-time data for temporary monitoring or from the analysis of data and view of a particular request. When any kind of data is requested, the query execution is linked with the storage component where data modelling has been done and data is stored in order to properly organize data. The storage involves the indexing and metadata management for easy and fast retrieval of data and analytics.

5.5. Testing the usability of the artifact

The usability of the proposed prototype has been tested and evaluated in two ways, first is by using the experts' suggestions mentioned in the questionnaire responses, secondly by support of the literature review by analyzing the shortcomings and enhancement suggestion by researchers about already existing models.

5.5.1. Evaluation Process using Questionnaire Responses/ Suggestions

The results from the questionnaires depict the importance of an enhanced CDM model for IoT to work for the improvement of adaptability and easy data management which should be both time and cost effective. IoT systems are capable of generating huge bulks of data on routine basis. It is of vital concern that a data management model should be devised to for efficient handling of huge amount of data being produced by IoT systems and devices.

In order to handle the huge bulks of data from various heterogeneous sources, advanced techniques have been applied to clean and process data effectively. The respondents identified that the enhanced system must be able to handle data security, integrity, scalability, availability, compatibility, transformation, and transportation. In the proposed model, the researchers have tried the best possible to cater with these user requirements. Latest and cost-effective techniques have been implemented. To lower the hardware cost, multiple low level servers have been used to store data nodes and resources sharing has been implemented. Similarly, in order to make the system fault-tolerant, data has been stored with replication of data nodes and easy access through indexing and meta-data management.

Data is cleaned by splitting, streaming and processed in parallel manner for easy availability, transformation ad mobility. The architecture of proposed cloud data management model warrants an active interaction between the front and back end of the IoT systems infrastructure. It serves as the basis of communication mechanism between IoT components with query propagation and results generation in an efficient manner with various connected modules as depicted by the lifecycle model. It enhances the storage capability of cloud computing to support resourceful scalability, data generation, integration, analysis and execution. The model of proposed data management mechanism targets the proper data migration to the cloud for storage and further analysis and execution of queries and process for IoT devices and systems.

Evaluation Process using Literature review

The research and study of the already existing literature of scholars has helped a lot in identifying the limitations of the existing cloud data management models and to propose an enhanced model with distinguishable and adaptable features. The data management architecture being proposed ensures a dynamic interactive mechanism between the front and back end of the computing infrastructure of IoT systems. The data management according to (Abu-elkheir et al., 2013) serves as the communication platform of an IoT system and also involves the propagation of query commands and results generation to and from various connected sensors and smart objects. The proposed data management model and architecture integrates the storage capability of cloud computing to IoT to support and manage data generation, analysis and executions. The discussed life cycle model serves as a benchmark towards guiding proper data migration to the cloud for storage and further analysis and executions for IoT systems and devices.

The proposed data management lifecycle model is formulated through tailoring of (Ma et al., 2013) data management reference model for IoT. This process is necessitated because the technological paradigm of IoT is still at its infancy state hence possess different challenges which must be properly addressed. To properly address the impending difficulties in data management process of IoT systems and devices, the proposed model seeks to serve as a guideline towards ensuring that proper IoT application and data management and also as a future reference for data management research in IoT data management.

At the data collection and aggregation module, sensors as well as smart objects data are collected at time intervals. This data can be collected at various network points or gateways for filtration and analysis. Since the IoT is characterized as a technological paradigm which constitutes wireless communication technologies, technological computing devices such as WiFi, RFID, Zigbee, and cellular infrastructures constitute major components used in data transmission that is sending, receiving and collecting of data. These technologies are known to transmit in real-time and as such according to (Abu-elkheir et al., 2013) due to bandwidth limitations and expensive data streaming, raw data transmission over the network is typically avoided. Due to this fact from the aforementioned researchers, it is pertinent that techniques or methods are deployed for summarization and also for merging operations in real-time to increase efficiency in data management of IoT systems and also compress the amount of data collected for storage. Hence, big data storage archives and compression have been implemented to cater the said need.

Following (Abu-elkheir et al., 2013) research, margin for data uncertainties needs to be catered at data reprocessing and cleaning phase in order to avoid a fatal error that may be caused later due to any discrepancies left in raw form from any of the data sources.

(Ma et al., 2013) states that the unreliability of data produced by IoT smart objects are referred to as unsophisticated data and are usually categorized into four major forms: False Negative, Invalid, Redundancy and False Positive. The false positive is categorized as noise; this noise is caused due to sensors itself or due to their operational environment. False negative is described as the data loss through ioT sensing devices, this issue generally results in signal interference between sensors or smart devices (Ma et al., 2013). Hence data cleaning and reprocessing constitutes an important component of the system, and uses temporal granule and spatial granule approach for the purpose. (Ma et al., 2013) study describes that albeit the it is important to ensure the data completeness and dynamics simultaneously. Therefore, both the low-level technique (Temporal granule) and high level (Spatial granule) techniques have been implemented to fulfil the requirement.

According to (Abu-elkheir et al., 2013), data can be kept dispersed at multiple storage devices or multiple servers but it is necessary to use a centralized approach for easy access however storage capacity needs to be calculated in order to attain scalability. This has been done by using a centralized approach of resources sharing in the proposed system. Similarly, data processing is necessary due to the need to properly interpret, manipulate and transmit useful data for IoT, this is to ensure proper service delivery in operational functions, it also entails that all data supplied to various components of an IoT systems are correct and also relevant to the operations. This process also needs appropriate techniques or technologies to ensure proper classification of data commands into hierarchical sequence of importance.

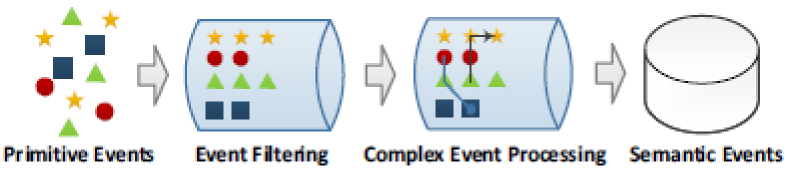

Figure 8: Complex Event Processing in IoT (Ma et al, 2013)

Based on (Ma et al., 2013) suggestions, a process in IoT is simply an occurrence of interest in a certain time, and this can be either referred to as primitive or complex events. In the paradigm of IoT, the processed or cleaned data can be regarded as primitive event. Complex semantic events are usually taken from low-level primitive events in a bid to aid the higher-level applications also referred to as a complex event processing (CEP) as depicted in Fig. 8. Hence, event process techniques have been implemented and discussed to make data ready for storage at data services module.

According to (Abu-elkheir et al., 2013), data query constitute a vital role in data management and may comprise of requesting of real-time data or analysis purposes, so both of the requirements have been made available in the proposed model for localized data and globalized view of data with insights of proper trends and patterns.

5.6. Conclusion

The discussed data management reference architecture represents a data management reference model suggesting the process needed to ensure data reliability and accuracy. The overall architecture of the enhanced Cloud Data Management model for IoT has been proposed with explanation of various modules in detail, starting with the general architecture of cloud data management paradigm, and the description of lifecycle model. The lifecycle model is a systematic representation of a proposed flow of data in an IoT system, this flow is necessitated in order to address data management issues as discussed in the literature review in this research. Furthermore, the usability of the system has been tested and justified through identification of limitations of existing models and description if possible solutions implemented catering the users' requirements and researchers' references have been explained.

5.0. Conclusion and Recommendations

The adoption of technological changes in the Cloud Data Management system is not a new thing, but especially when it comes to the unpredictable bulks of data flowing around the global network. This flow of data is supposed to be increasing day by day in a probably uncontrollable fashion. Hence, there are many hurdles and challenges in the implementation of Cloud Data Management systems for Internet of Things, that need to be tackled one by one in the attempt of making the CDM models better to cater the emerging needs. The active use of IoT devices in all the sectors of real life give a strong hope that the proposed system can be easily absorbed and merged into the current infrastructure, and no new or special expensive devices need to be brought in to apply the techniques of this system.

Moreover, the researchers believe that the proposed system can be enhanced in a more efficient manner and using the latest technologies for time and cost effective and a long-term solution for the revolutionizing Internet of Things data around the globe and devices. The researchers suggest that the system should be extended in a manner that clients should be given some system authentications to cater their basic and emergency needs as soon as they experience some issues on the cloud server for their data. This will enable end-users and clients to back up their data and work on it as convenient.

Also, it is very necessary to interview both the major big data clients and hosting companies before initiating the task of extension, and efforts should be made to improve current version of the system by incorporation of the solutions of shortcomings as highlighted by the technical experts.

6.0. References

Alzahrani, A., Alalwan, N. & Sarrab, M. (2014). Mobile cloud computing: advantage, disadvantage and open challenge. *Proceedings of the 7th Euro American Conference on Telematics and Information Systems - EATIS '14*. [Online]. p.pp. 1–4. Available from: http://0dl.acm.org.cataleg.uoc.edu/citation.cfm?id=2590651.2590670\nhttp://dl.acm.org/citation.cfm?doid=2590651.2590670.

Angarita, R. (2015). *Responsible Objects: Towards Self-Healing Internet of Things Applications*. [Online]. p.pp. 307–312. Available from: http://ieeexplore.ieee.org/ielx7/7264388/7266915/07266985.pdf?tp=&arnumber=7266985&isnumber=7266915\nhttp://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=7266985.

Abu-elkheir, M., Hayajneh, M. & Ali, N.A. (2013). *Data Management for the Internet of Things: Design Primitives and Solution*. p.pp. 15582–15612.

Belli, L. (2015). *Big Stream Cloud Architecture for the Internet of Things*. p.pp. 5–6.

Bonomi, F., Milito, R., Zhu, J. & Addepalli, S. (2012). Fog Computing and Its Role in the Internet of Things. *Proceedings of the first edition of the MCC workshop on Mobile cloud computing*. [Online]. p.pp. 13–16. Available from: http://doi.acm.org/10.1145/2342509.2342513\npapers2://publication/doi/10.1145/2342509.2342513.

Fan, T. & Chen, Y. (2010). A scheme of data management in the Internet of Things. *2010 2nd IEEE InternationalConference on Network Infrastructure and Digital Content*. [Online]. p.pp. 110–114. Available from: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5657908.

Gang, C. (2010). Data Center Management Plan in Cloud Computing Environment. *2010 3rd International Conference on Information Management, Innovation Management and Industrial Engineering*. 4. p.pp. 393–396.

Gubbi, J., Buyya, R., Marusic, S. & Palaniswami, M. (2013). Internet of Things (IoT): A vision, architectural elements, and future directions. *Future Generation Computer Systems*. [Online]. 29 (7). p.pp. 1645–1660. Available from: http://dx.doi.org/10.1016/j.future.2013.01.010.

Ji, C., Li, Y., Qiu, W., Awada, U. & Li, K. (2012). Big data processing in cloud computing environments. *Proceedings of the 2012 International Symposium on Pervasive Systems, Algorithms, and Networks, I-SPAN 2012*. [Online]. p.pp. 17–23. Available from: http://www.scopus.com/inward/record.url?eid=2-s2.0-84874615486&partnerID=40&md5=ae7546b5e769b1735c7c75f35ac99df7.

Krensky, P. (2015). *DATA MANAGEMENT FOR THE INTERNET OF THINGS*. [Online]. Available from: www.aberdeen.com.

Kshetri, N. (2012). Cloud computing in developing economies. *IEEE Computer*. [Online]. 43 (10 April). p.pp. 47–55. Available from: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2015387.

Lee, I. & Lee, K. (2015). The Internet of Things (IoT): Applications, investments, and challenges for enterprises. *Business Horizons*. [Online]. 58 (4). p.pp. 431–440. Available from: http://dx.doi.org/10.1016/j.bushor.2015.03.008.

Ma, M., Wang, P. & Chu, C.-H. (2013). Data Management for Internet of Things: Challenges, Approaches and Opportunities. *2013 IEEE International Conference on Green Computing and Communications and IEEE Internet of Things and IEEE*

Cyber, Physical and Social Computing. [Online]. p.pp. 1144–1151. Available from: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6682212.

Petersen, K. & Gencel, C. (2013). Worldviews, research methods, and their relationship to validity in empirical software engineering research. *Proceedings - Joint Conference of the 23rd International Workshop on Software Measurement and the 8th International Conference on Software Process and Product Measurement, IWSM-MENSURA 2013*. p.pp.81–89.

Rui, J. & Danpeng, S. (2015). Architecture Design of the Internet of Things Based on Cloud Computing. *2015 Seventh International Conference on Measuring Technology and Mechatronics Automation*. [Online]. p.pp. 206–209. Available from: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=7263548.

Ruoss, S. (2016). *The Cloud is here – now comes the Fog*. [Online]. Available from: http://web.fhnw.ch/projekte/cloud-days/cloud-use-cases-days/3-cloud-use-cases-day-2016/pdf-usecases-2016/10-cisco-cloudfog.

Saharan, K.P. (2015). Fog in Comparison to Cloud : A Survey. *International Journal of Computer Applications*. 122 (3). p.pp. 10–12.

Smith, A., Bhogal, J. & Sharma, M. (2014). *Cloud computing : adoption considerations for business and education Review of Cloud Computing*.

Thomas, M.O. & Rad, B.B. (2017). *Reliability Evaluation Metrics for Internet of Things , Car Tracking System : A Review*. (February). p.pp. 1–10.

Zeng, L., Veeravalli, B. & Wang, Y. (2010). *Cloud Data Management*.

Appendix 1

Questionnaire Design

Questionnaire designed check about the credibility and importance of a Cloud Data Management model:

Section A: General Information

Q1. Please specify your role in organization:

- ☐ CEO/ CTO
- ☐ Project Manager/ Coordinator
- ☐ Consultant
- ☐ Data Scientist
- ☐ Business Analyst
- ☐ Other

Q2. Please specify your Gender:

- ☐ Male
- ☐ Female

Q5. What is your experience level?

- ☐ Less than 5 years
- ☐ 5-10 years
- ☐ More than 10 years

Q4. What is your qualification?

- ☐ Degree
- ☐ Masters or Above

Section B: Cloud Data Management Model

Q1 What do you know about the Cloud Data Management Model?

____

Q2 In your opinion, how does a Cloud Data Management Model assist the organizations to carry out and perform their business processes?

Q3 Is Cloud Data Management Model being used in your workplace? If yes, please elaborate the process.

Q4 If the answer to the previous question is yes, kindly let us know which hardware is being deployed to make use of this technology?

Q5 What difficulties are being faced by you, if any, in using the Cloud Data Management technology?

Q6 Do you think that a Cloud Data Management Model can play an important role in today's world when used for the Internet of Things?

Q7 What benefits does the use of Cloud Data Management Model for the Internet of Things hold for the end-users/clients?

Q8 In your opinion, what measurements should be taken in account to improve already existing CDM models for future needs of IoT?