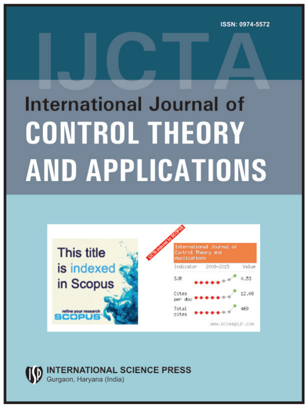

International Journal of Control Theory and Applications

ISSN : 0974-5572

Volume 9 • Number 43 • 2016

The Hype of Emerging Technologies: Big Data as a Service

Babak Bashari Rad[a], Pouya Ataei[b], Yasaman Khakbiz[c] and Nafisseh Akbarzadeh[d]

[a-c]School of Computing, Asia Pacific University of Technology and Innovation (APU), Technology Park Malaysia, Bukit Jalil, Kuala Lumpur, Malaysia. Email: [a]babak.basharirad@apu.edu.my; [b]pouya.ataei.7@gmail.com; [c]yaskhakbiz@gmail.com
[d]School of Technology, Asia Pacific University of Technology and Innovation (APU), Technology Park Malaysia, Bukit Jalil, Kuala Lumpur, Malaysia. Email: nafiseh.akbarzadeh@gmail.com

Abstract: Technologies comprise more aspects of modern societies day by day. As a result, a great blast of data has been perceived in global scale. Big data impacted many industries ranging from healthcare to transportation, agriculture, forestry, and even government. Big data can provide insights, solutions, and opportunities in all sort of businesses. However, a great number of Big Data projects are facing a big failure. This paper aims to explore various facets of big data and big data as a service, as well as, the reasons and challenges behind the vast amount of failures in big data projects with a focus on origin of these problems.

Keywords: Big Data, Big Data as a Service, BDaaS, Cloud, Analytics.

1. INTRODUCTION

The value and benefits of Big Data is no longer unrecognized. According to the most recent GRIT report [1], big data is an important issue in today's marketing research due to the opportunities and challenges it delivers. Traditional "goods-dominant" business approaches are being overridden by services and technologies in the market industry. In Gartner's report 2014, big data was announced as a hype circle of emerging technologies [2]. Furthermore, in regards to a CIO survey [3], 55% of big data projects around the world were not successfully finalised. This paper aims to present an overview of the nature of big data as well as big data services. Additionally, it will discuss why adapting big data as a service is challenging. The rest of this paper is classified as per the following: section 1 provides an overview and better understanding of this paper; Section 2 is dedicated to big data as a service; Section 3 discusses the essential characteristics of big data as a service; Section 4 is about data models, from traditional data models to modern and neo metropolitan data models; Section 5 delves into data knowledge models; Section 6, context driven knowledge models; 7 provides data paradigms; next is section 8 about evolvement of big data; 9 represents choke points; section 10 is the conclusion and in the last section, list of references is provided.

2. BACKGROUND STUDY

A. The Era of Big Data

Big Data and the Internet of Things together are driving the third industrial revolution in which raw elements are calculated and measured in bites [4]. We are living in an age of Big Data where according to IDC (International Data Corporation) the big data market scale will be 20 billion USD until 2016 and the big data market growth rate will be nearly 25% [5]. Before the broad rise of the Internet, data was being accumulated only by employees of companies, but now this has been changed, individual users or even machines and tools are able to generate colossal amounts of their own data per day. This large diverse volume of information is called "Big Data". Some of the main big data generators are: Sensors, Social networks, Mobile devices, Internet transactions, and service-generated data such as trace logs, Quality-of-Service (QoS) information, service invocation relationship, etc. [6]. Big data is usually described by five characteristics, the "Five V's": Volume of data which refers to data size; Velocity, which is the speed of incoming data; Variety, termed as, diverse types of data; Veracity is accuracy of data and finally Value, is the added-value of big data in order to help companies recognize their client's demands [6]. Big Data cannot be processed with traditional approaches because of their limitations; and to solve this issue, an algorithm called MapReduce has been represented. MapReduce is a data processing model used for large clusters of data [7]. Hadoop is one of the most famous open source frameworks using MapReduce algorithm where data is processed in parallel on distributed datasets across computer clusters. According to IDC [5], world store of data is now counted in zettabytes. That is trillions of gigabytes and 90 percent of this amount has been generated in the last two years. It is very obvious that the traditional approaches are now out of use. Cloud data storage is a new technology, which is beneficial for addressing Big Data storage issues. A cloud database is a cloud platform as a service that allows users to manage, store and retrieve data from the cloud. Services using the Hadoop platform in conjunction with cloud based databases, for the purpose of analysing, managing and storing Big Data, is known as, Big Data as a Service or BDaaS [8]. Figure 1 visualizes a brief overview of the big data revolution.

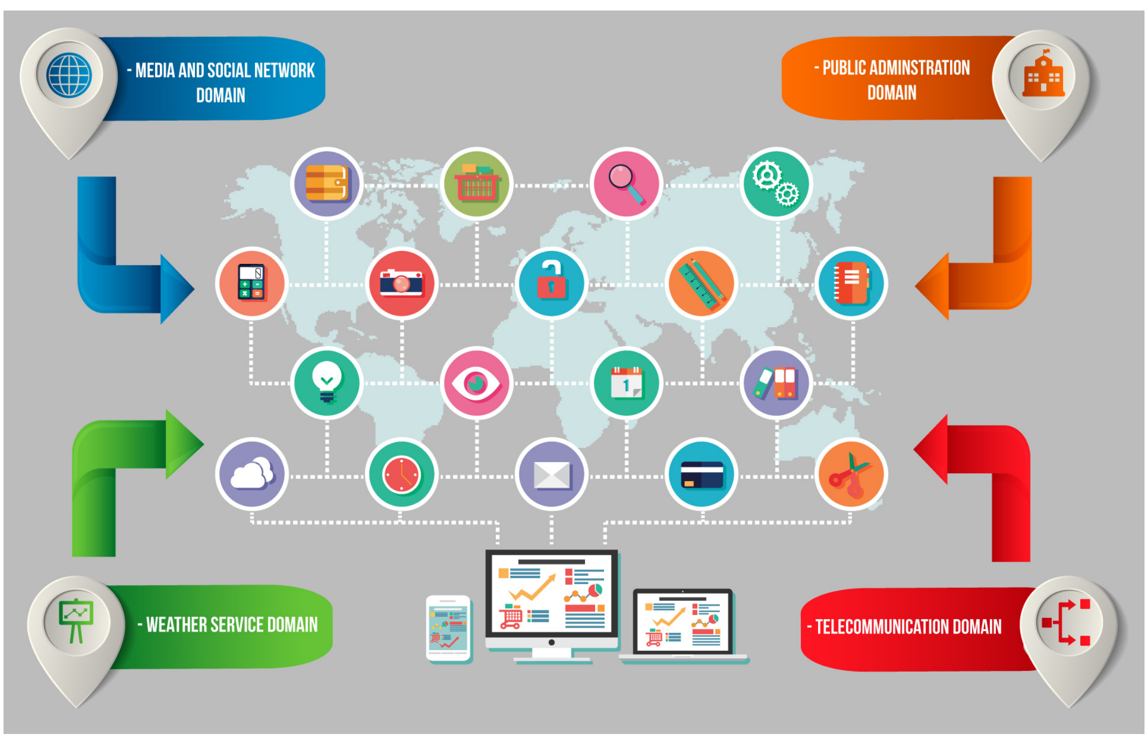

Figure 1: Big Data Domains

B. Applications of Big Data

Big Data has attained considerable attention in diverse environments such as healthcare, government, and academia. This paper provides some examples of how big data is being utilized in today's world to give an overview of how vast and inclusive this subject can get.

- The former National Security Agency (NSA) subcontractor, Edward Snowden who leaked some top secret information about how NSA conducted surveillance activities on US citizens and other parties made an obvious example of how governments make use of Big Data for their own purposes [9].
- Nowadays political campaigns leverage the power of big data to expand connectivity between their supporters using social media to generate more votes as well as more donations [10].
- Wall Street analysts are progressively using data from Internet search trends and social media in their investment strategies. The way we are trading the world is being enhanced by big data. Surveys show that approximately 50 percent of investment firms are using big data [11].
- In Transport facilities, the use of big data can be seen quite clearly these days, new applications such as Uber or Waze are taking advantage of big data in order to offer information from different unique sources into a real time database [11].
- Benefits of big data in healthcare can be categorized in 3 areas: identify risk factors, prevent disease, and design intervention for health behaviour change [12].
- The value of big data in agriculture is now being recognized. Big data helps in monitoring the demand of users as well as optimizing the strategies for better productions [13].
- Big data technologies are now being used to improve tourism activities. Many tourist companies around the world are using this technology to get valuable insights like better understanding of tourist behaviours, needs, and preferences [14].

C. Big Data Challenges

The large growing volumes of big data are causing different challenges. The main challenge is to deal with the complex nature of big data, which are the 5 "V's".

1. *Volume:* Volume of data refers to how big the data is. According to Yousefi and Chiadmi [15] regarding to Volume of Big Data in 2013, Facebook generates 350 GB in every minute and twitter 277,000 of tweets. Moreover, according to the same research, google users execute two millions search queries per minute. In general, supercomputing industry is more focused on volume aspect of big data.
2. *Velocity:* Velocity represents the concept of how fast the data is processing and storing. Researchers in Internet of Things and sensor web areas are most likely interested on velocity aspect of big data.
3. *Veracity:* Veracity is an integral segment, by the reason that, the time, during which the data can be processed and analysed is a crucial matter. IBM added Veracity to the other core three "V's" that refers to the ambivalence of data and brings a challenge of how to trust the accuracy of information [16]. Most likely, social scientists and humanities are concerned with the veracity front of big data.
4. *Variety:* The fourth "V" is Variety. Clearly, there are different types of data such as sensors, pictures, videos, texts, etc., that are categorized under two major families, called structured and unstructured data.
5. *Value:* Value is the fifth characteristic, which is the added-value of big data so as to help companies to acknowledge their customers' requests. Figure 2 illustrates the five characteristics of big data.

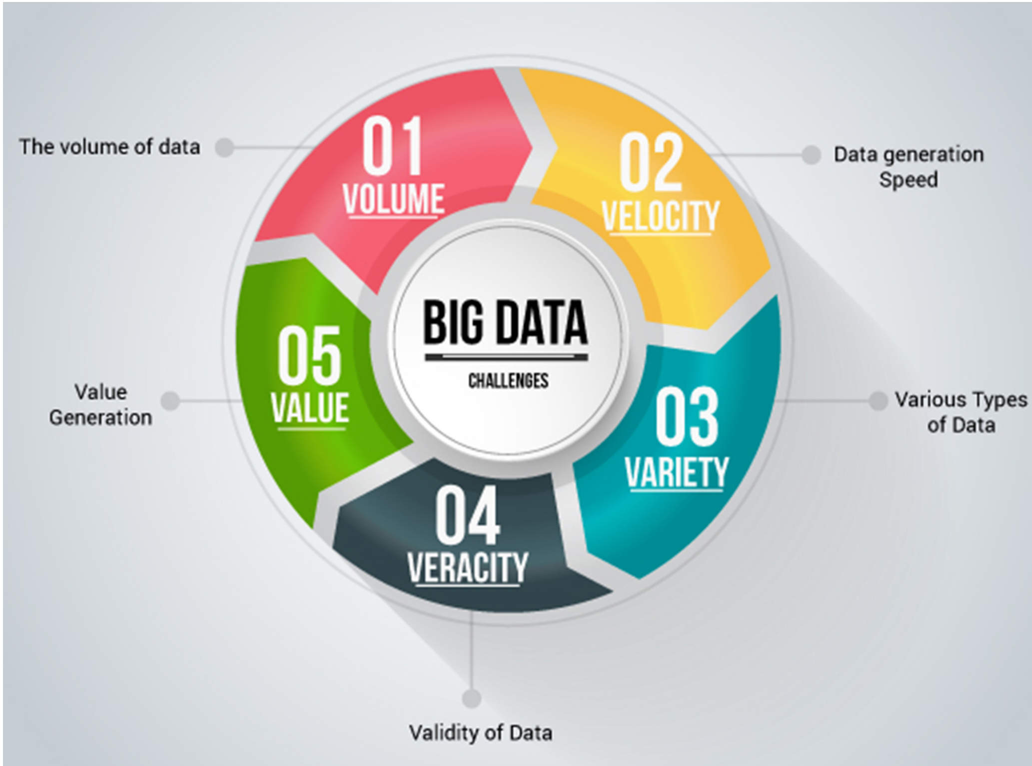

Figure 2: The Five V's

D. Open Source Platforms

As result of solving these challenges, many open source Big Data frameworks and platforms are available. Several MapReduce Frameworks such as Skynet, FileMap, Apache Hadoop and Sailfish are now being utilized, among which, Hadoop is the most famous and prominent. These frameworks are efficient for many use cases. They are based on Massive Parallel Processing (MPP) and distributed storage of data; mostly these systems are using Master-Slave architecture and data replication to ensure no trace of failure [12].

3. BIG-DATA-AS-A-SERVICE

Big-Data-as-s-Service (BDaaS) is an architecture, which represents common big data services that are provided as cloud hosted services [8]. This means Cloud computing is a nature of Big Data architecture. Although, other distributed architecture may be used to host the services. In general, BDaaS provides diverse types of big data services and analytics. According to Zheng et. al., [6], BDaaS encloses Big Data technologies into three different layers as follows: infrastructure-as-a-service (BDIaaS), platform-as-a-service (BDPaaS) and software-as-a-service (BDSaaS). Figure 3 illustrates an overview of BDaaS.

A. Infrastructure-as-a-Service

IaaS also known as hardware as a service, is a service model, which supplies computer infrastructure on an outsourced basis including storage, hardware, servers and datacentre space [17]. IaaS allows companies to use public cloud providers' infrastructures. Mainly, IaaS provides storage capacity and computing processes for big data, using Storage-as-a-Service and Computing-as-a-service facilities. The biggest challenge of IaaS is the requirement of supporting many different types of data (variety).

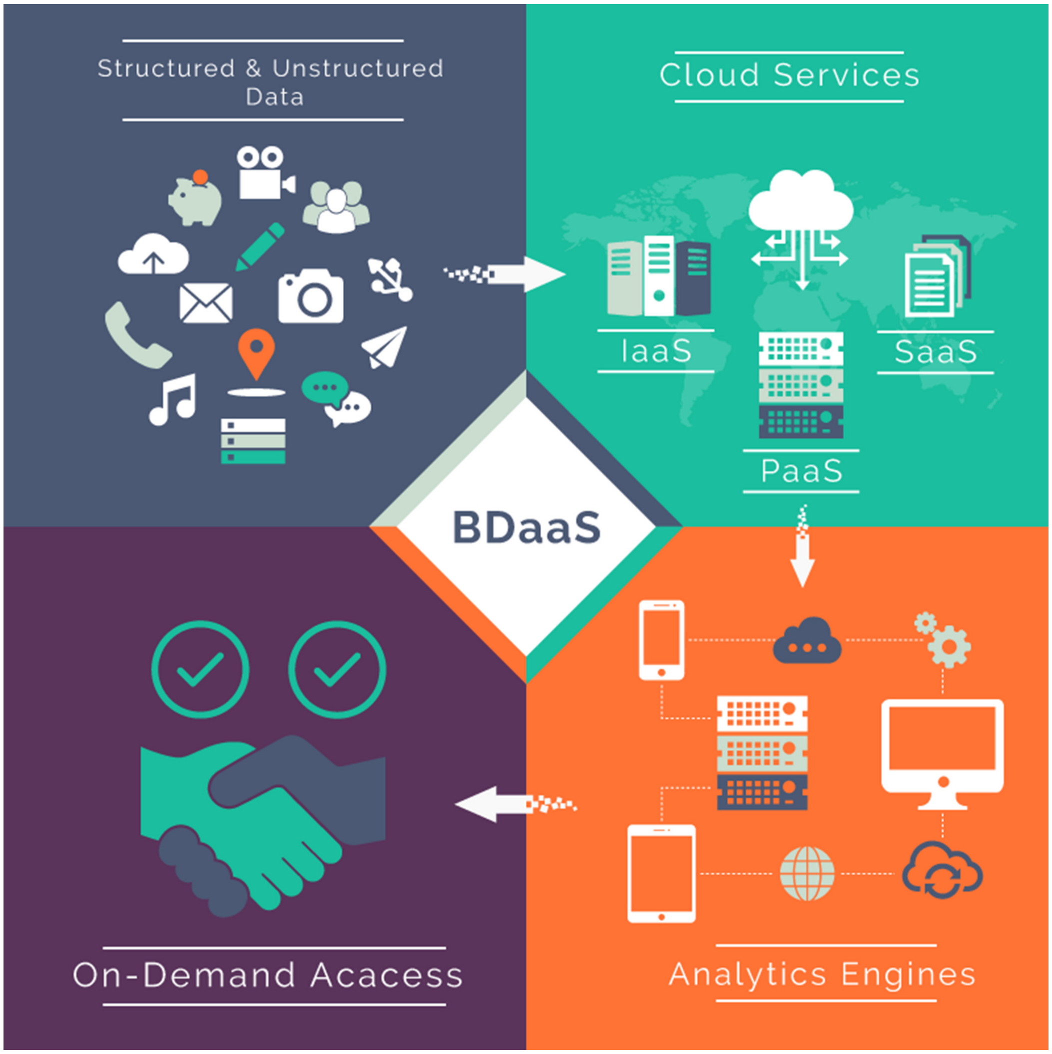

Figure 3: BDaaS in a Nutshell

B. Platform-as-a-Service

According to EMC solution group [18], PaaS describes a computing platform on the Internet, which is delivered or rented as a service. PaaS is in between IaaS and SaaS in cloud computing layers. Clients including software and hardware plus application (SaaS) are above PaaS. Server including hardware and software and IaaS are below PaaS, in five levels of cloud computing. PaaS can be leveraged in public or private cloud environment to access, analyse, design, deploy, and implement services and applications on large data sets. Google's BigQuery is a good example of PaaS, which allows users to analyse Big Data and take advantage of real-time business insights.

C. Software-as-a-Service

SaaS uses the Internet to host software applications. This service allows users to perform self-service analysis, provisioning, and collaboration. Mostly SaaS vendors provide two principal facilities: it updates the applications automatically and costumers do not need to provide any hardware for this matter [17]. In recent years, the use of SaaS has become mainstream. SaaS is used in many business areas such as customer relationship management.

D. BDaaS Types

Referencing to these layers, BDaaS can be divided in into four types:

1. *Core BDaaS*: Core BDaaS is rather inclusive, it makes use of basic infrastructure's alike Map Reduce, Hadoop and Spark [19]. Hadoop is very popular among users as it is an open source software. This allows users to alter code and have access to a greater knowledge base of support. For storage purposes, Core BDaaS makes use of applications like Amazon's S3 or NoSQL, and Hive. An example for complete Core BDaaS is Amazon's Elastic MapReduce (EMR).
2. *Performance BDaaS*: According to Prokopp [19], performance BDaaS uses the basic infrastructures similar to Core BDaaS with the difference of making usage of different hardware and software services to elevate performance and increase scalability.
3. *Feature BDaaS*: Feature BDaaS in another approach of BDaaS, which has been evolved to comprise features beyond the usual Hadoop ecosystem aims [20]. This feature involve programming and web interfaces, as well as database adaptors.
4. *Integrated BDaaS*: According to Prokkop [19], this approach combines Feature BDaaS and Performance BDaaS in order to extend supporting and performance. Albeit, it should be mentioned that this approach is still theoretical, and practically none of the services offer it, by this moment. Figure 4 illustrates the four BDaaS types.

E. BDaaS Framework

There are six different layers in BDaaS framework according to their functionality in the procedure of analysis, data storage, and computing.

1. *Data Infrastructure*: The core base of BDaaS consists of several nodes and data hardware. This layer provides backup systems to counterbalance data safety in the network. Today, these infrastructures are the best solution, considering how expensive it is, to establish dedicated infrastructure.
2. *Cloud Infrastructure*: Cloud infrastructure is where data, hardware, and software can get related together in a virtualized domain. Cloud Infrastructure can be reserved on demand, on spot, or for a period of several years in advance. It can also be public or private [21].

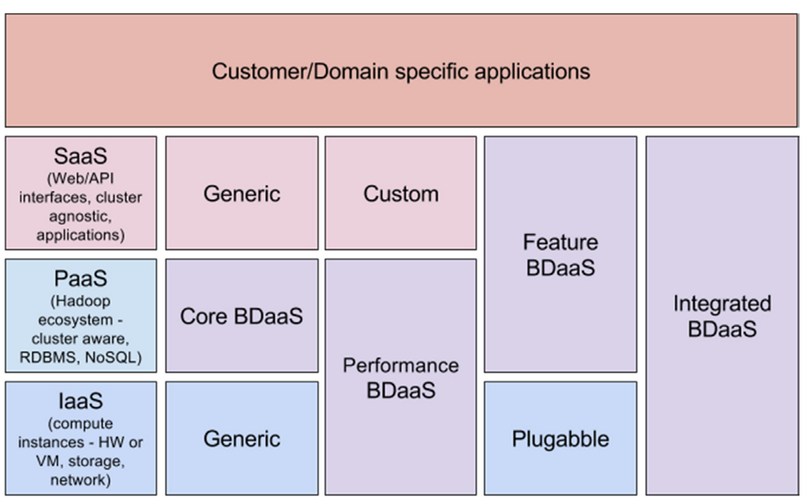

Figure 4: The four types of big data as a service [19]

3. *Data Storage Layer*: This layer is highly scalable; it allows the introduction of new nodes to the network for costumer's requirements of velocity, volume or variety. This layer is accessible for costumers.
4. *Computation Layer*: The aim of this layer is to manipulate and manage data to fulfil the needs of customers. It consists of framework processors and APIs to support this objective.
5. *Data Management*: This layer is responsible to make sure that data is secured and efficient. For this purpose, this layer undertakes resource requirements, system backups, and deployments over the cloud.
6. *Data Analysis*: The purpose of this layer is to analyse underlying data procedures. This layer provides graphical tools and wizards through the process of accessing data. Data Analysis layer also offers customized path for the individual industry base needs of customers. Figure below illustrates BDaaS framework with all corresponding layers.

F. Big Service: The Essential Characteristics

BDaaS is an umbrella term to portray a wide variety of outsourcing of big data services in the cloud [21]. As aforementioned, there are three cloud service models by the NIST standards: Infrastructure-As-A-Service (IaaS), Platform-As-A-Service (PaaS), and Software-As-A-Service (SaaS) which together offers six essential characteristics, as following;

- On-demand
- Self-service
- Broad network access
- Resource pooling
- Rapid elasticity
- Measured service

In what follows, these characteristics will be outlined in details.

- *On-demand:* the on-demand model provides a means to grant clients with resources, as they desire. The model has been merely designed to respond to volatile requests that different corporations may have. As enterprises may demand on computing resources on variable basis (for example it may be drastically escalated during peak hours), the Cloud service provider has to maintain sufficient resources to meet the fluctuating demand. On-demand computing is one of the most pervasive enterprise computing models, and it serves as an integral segment of BDaaS.
- *Self-service:* self-service computing, as the name states, is a private cloud service in which, clients are allowed to provision storage and launch applications without going through an external cloud service provider. As in the case of BDaaS, benefits include the user being able to access storage, applications, and other services without relying on their internal provider handling it. Everything is automated to lift burden off the IT department.
- *Broad Network Access:* this refers to mechanisms that allows access to resources hosted in a private cloud network (operated within a company's firewall) that are available for access, regardless of the device. Besides, these resources are also accessible from a wide variety of locations that offer online access. BDaaS, make use of this mechanism to eliminate access limitations of the clients.

- *Resource Pooling:* resource pooling is a Cloud structure that serves multiple clients, or tenants with provisional and scalable services. These services may adjust to suit each client's needs, without any changes being apparent to the end user. BDaaS, benefits from this mechanism, by creating a sense of infinite or immediately available resources by controlling resources adjustments at the meta-level.
- *Rapid elasticity:* is a cloud computing term for scalable providing. Complementary to resource pooling, rapid elasticity allows clients to request additional space automatically in the Cloud. This is a very essential aspect of BDaaS.
- *Measured Service:* in a measured service, aspects of the cloud services are controlled and monitored by the Cloud provider. This is momentous for billing, access control, resource optimization, and capacity planning. Measured service, is a pivotal factor for BDaaS architecture, as it controls the efficiency and overall flow of the provided services.

4. THE SHAPE-SHIFT: SYSTEM DEVELOPMENT AND INFORMATION PROVISIONING

System development and information processing, as a means to account for faster and smarter works by machines, have gone through distinct phases, in order to ensure a harmonious interaction with society and the way our species have been evolved and changed within last 60 years. It began with the emersion of Eniac in 1946, and commenced to become increasingly sophisticated on every major phase of its growth. From the era of mainframes to distributed processing, networking, internet, and lately mobile, there has been major contributions and development toward addressing our needs in an increased engineered and efficient manner. Not only the protocols, technologies and techniques of software development and information processing has been evolved, but the conceptual framework and the way in which, the human and computer interacts has been revolutionized too. In what follows, we will discuss three major models of system development and information processing with detailed attention to how new models of conduction has been derived from the prior, and what were the major role players in this transformation.

A. Traditional Model

Traditional models were defined at the very beginning of the computer science. Software engineers architected and initialized a sequenced activities in order to follow up with the intended requirements, and to deliver the quality end products. Hence, in achieving the end product, the involving activities were being controlled, underlying a rigorous discipline. These disciplines were involved heavily in statement of detailed stages. Besides, the stages themselves were composed of complex and colossal documentations. These bureaucratic documentations were practicing high level of details in the chosen methodology, like Waterfall, which required a lot of time to apply. Hence, the work of managing the methodology itself was sometimes more than the work on the intended system. These old good-dominant methodologies, despite being effective at the time, did not last for long. By the emergence of networking, and internet, the perspective of the whole industry has shifted from seeing customers as an 'uninvolved recipients of goods' to the 'cooperators' of generating value. This implied more than just a move from goods to services; it also refactored the purpose and overall approach of enterprise and its value creation procedure [22].

Ensuing that, organizations started to revise their perspective toward the relationship with their customers profoundly, and the level of involvement and interaction of the client. This shift in perspective paved the inception of a new era, an era of rapid application development and crowdsourcing.

B. Metropolitan Model

Customary, computer scientist, system analyst, and software engineers were primarily concerned with 'value generation' of firms, and not on 'value co-creation'. The closest approaches to 'co-production' were through methodologies such as Rapid Application Development, Joint Product Design, Joint Application Design, and in recent times, Agile methods that presented an iterative model of attaining immediate customer feedbacks along with the development process [22]. Yet, all these approaches are under the trajectory of 'good-dominant' philosophy.

While product-concentrated and goods-concentrated approaches treats clients as secluded entities, service-oriented approaches, on the contrary, emphasizes on assimilation of resources from various entities, including customers, firms, suppliers, networks, and any other stakeholder that can play a role in value co-creation.

The emergence of these approaches is currently dominating the industry, as huge role players such as Facebook, Amazon's Mechanical Turk, Wikipedia, and many other service-based public systems have been arose underlying the same philosophy. Traditional models were underpinned by 'closed world' assumptions, in which, projects have dedicated pre-specified resources, requirements are acknowledged, and systems are designed, developed, tested, and deployed in planned phases. Nonetheless, these assumptions all break down in the face of current crowdsourced and service-oriented world. Metropolis, the Greek word for "city", as a new model to comply with modern world trends and requirements, were therefore, proposed by Kazman and Chen [22]. According to Kazman and Chen [22], the new model of producing systems is more like constructing a city than a single building, a perspective called 'ultra-large-scale (ULS)'. In this perspective, systems are not limited to be built by a single organization with centralized control, and are involved with global volunteers for continuous evolution and value co-creation. Metropolis model, with inherent support for crowdsourcing, and service-oriented architecture, were underpinned by their application to massive and rapidly growing set of software-centric systems. As demonstrated by Kazman and Chen [22], benefits of the unified wisdom, ingenuity, and capacity of the congregations, has made Metropolis model prevalent to the traditional ones, which in return inevitably led business and project managers to approach variously about disparate aspects of system development including design, implementation, requirement elicitation, architecture, testing, delivery, and maintenance.

C. Neo-Metropolitan Model

Neo-Metropolis, predicated on Metropolis model, proposed in 2016, by the founders of Metropolis model, Kazman and Chen [22], plus a few more collaborates, to provide a viable alternative to circumvent obstacles in implementing a big data strategy. Neo-Metropolis, is an implementation model for a service-oriented platform development, such as BDaaS (big data as a service). Big data systems, being at hype of emerging technologies, have received an enormous attention lately, which has brought up a lot of researches regarding its different aspects and implementations. According to Xu et. al., [23], Giant IT companies such as Facebook, Microsoft, Apple, Oracle, and SAP have invested more than 15$ billion buying big data management and analytics software firms other initiatives to solidify their foothold in this field. Traditional approaches were not effective any more to process an overwhelming service-generated data, and Metropolis model were not exclusively accentuating Big Data, thus, Neo-Metropolis came to play. As explained by Kazman and Chen [22], Neo-Metropolis model, despite sharing the basic tenets of Metropolis model, has some distinct features, such as offering an organized, coherent set of open-world innovation opportunities for vendors as well as for the platform's edge customers. Neo-Metropolis emphasizes on service-engineering and big data in enterprises in specific. The model depicts on how the Neo-Metropolis BDaaS can be developed and managed to foster innovation and productivity in a cloud platform that facilitates edge-customer innovation. The model has been derived from the analysis through

Information Systems (IS), and Software Engineering (SE) theories and practices that illuminates the characteristics of service-oriented architecture, cloud platform service management and development, and BDaaS.

Figure 6 depicts the involving major factors contributed to the overall shapeshift of the system development and information provisioning paradigms.

Figure 6: System Development and Information Provisioning Shapeshift

5. CONTEXT DRIVEN KNOWLEDGE MODELS

Under the trajectory of big data, crowdsourcing, Cloud, and emanation of modern and noble models of system development, data analysis and business intelligence has been challenged and revolutionized too.

Subsequently, after the emergence of big data, and an increasingly service-oriented engineering trends, business intelligence and analytics, transmuted to catch up with the leading edge technologies. These concepts, being embodied in the centremost of BDaaS, have begun to face their transformation, which will be elaborated in what follows.

A. Business Intelligence and Analytics

Business intelligence and analytics (BI&A) has materialized as a significant field of study, reflecting the volume and impact of data-related problems to be rectified. Based on IBM tech trend report, 4,000 information technology professionals from 93 counties and 25 industries, determined business analytics as one of the major technology trends [10]. Asides, in a survey of the state of business analytics by Bloomberg Businessweek [24], 97 percent of companies with revenues exceeding 100$ million were found to use some form of business analytics. Alongside data gradual cheapness and ubiquity, analysis mechanisms, databases, machine learning, econometrics, statistics, and visualization has been augmented. Data and analysis have shed lights on various organizations, and initiated a consequential interests in BDaaS, which is a nexus of techniques, approaches, systems, technologies, services, infrastructures, platforms, practices, methodologies, and applications that analyse crucial business data and illuminates companies with patterns of business and market.

B. Knowledge Modeling

Knowledge modeling, serving as an integral part of this service, has aided the transformation of 'global push' to a 'full service approach'. This idea depicts the fact that, BDaaS environment is an opportunity for massive collaboration between people and the operation of intelligent virtual objects and devices. Knowledge modelling is a means for creating a computer interpretable model of knowledge or standards specifications about certain kind of service, product, or facility. This concept plays a critical role for the business intelligence and data analytic segment of BDaaS. As structured and unstructured data begin to flow over the Cloud infrastructure, platforms, software, and analytical engines, they will be taxonomised and segmented into various units. These unites underlying different factors such as availability, scalability, reliability, elasticity, and flexibility constitute the knowledge model. These knowledge models, being each defined underlying their own unique crunching algorithms, software, mechanisms, techniques, and environments, therefore, generate a sense of 'context-driven' knowledge service delivery. This service in the Cloud, involves in retrieval and enrichment of knowledge content to be fully driven by consumer's needs and requirements. Preceding endeavours to conceptualize context-driven knowledge models were involved with deployment of digital knowledge workers. These knowledge workers were service-oriented programs capable of answering knowledge related queries via a knowledge server. These workers could be conceived as on-demand knowledge applications, which were capable of searching, analysing, and forming its knowledge resources [25]. A service engineered knowledge modelling, compared to traditional storage and tools oriented knowledge systems, has superior advantage, by expanding toward the offering and availability of on-demand knowledge through a unique context. Ju and Shen [26], emphasized that the development of knowledge service contributes to the development of workers as well as to the improvement of their productivity. The evolution of traditional knowledge management approaches (capitalization) towards service oriented architectures, equip companies with necessary knowledge to act upon and so, achieve their business goals, which further benefits by constituting the primary source of advantage; 'actionable knowledge'. These new dimensions illuminate the actionability of knowledge delivered to customers, and are positively supporting the systematic recognition of how consumers retrieve, interact, share, and contribute to enrich contents. On that account, context driven knowledge models need to "be conceived as interrelated frameworks of indicative factors in the management of intellectual capital".

6. DATA PARADIGM: INSTRUMENTALITY AND ITS ADVERSITY

Data being the rhythm of our current world's flow, despite being generated uncomplicatedly, is a very valuable source to any existing company looking for growth. However, utilizing these colossal amounts of structured and unstructured data does not go with ease. Since 1980s, numerous data mining algorithms has been developed by researchers from artificial intelligence, algorithm, BI & analytics, and database communities. Most of these foundational data mining algorithms have been incorporated in commercial and open source data mining systems [27]. Albeit, BDaaS, which involves in extensive interconnection and patterning for dynamic competitive advantage, behavioural psychographic target marketing, real-time decision making, and real-time brand risk management, requires an extensive orchestrated analytics. These complicated analytics, as aforementioned, are concentrated around five ever-increasing factors consisting of volume, variety, veracity, velocity, and value.

Volume depicts the ever-increasing amount of data streaming to analytics engine of BDaaS from various sources such as transactions, social medias, sensors, applications, e-commerce systems, etc. These colossal streams of data, provides data scientists with increased complexity in determining relevance, clustering, patterning, value creating, and deducing.

Variety of data nowadays, is not comparable to any other time in history. Today, data comes in distinct formats, from relational databases to distributed data stores produced by end users, to text documents, images, emails, meter collected data, and sensor collected data, videos, audios, stock ticker data, and financial transactions. This enormous variety of structured and unstructured data puts forward a heightened complexity for collection, clustering, analysis, and crunching.

Velocity According to Gartner report, velocity is both how fast data is being produced and how fast the data must be processed to meet demand [2]. This concept is the most challenging aspect of BDaaS, as the inherent distributed architecture of the Cloud may compromise on reliability, and that results in not providing in line with client demands.

Veracity is another aspect that provides with challenges. Any person with access to any digital devices, becomes capable of generating any sort of data, most of these may not be valid, or may be in poor quality, which later on leads to inaccurate patterns, and false decisions. For that sake, data has to be validated, and that is a complex process, consuming great amount of resources, both in terms of network and computational resources.

Value generation, being the core element of BDaaS, confronts analysis with intricacy. Colossal amount of data coming from multiple sources, underlying various knowledge models, analytic engines, validation rules, and velocity, makes a nexus of data that has to be scrutinized and evaluated. This is one of the biggest challenge, the IT world is facing.

7. EVOLVEMENT

From the era of the mainframes to BDaaS, human species have been evolved in number of major phases. The momentum began with traditional arid approaches to system development, and initiated the gradual shift from ‘good-dominant’ approaches to ‘service-oriented’ architectures. With emergence of crowdsourcing, and open source software, it was not companies that were developing software and systems no more, but the global community of developers federated to forge unprecedented solutions.

The ‘open-world’ approach removed the veil of conservativeness, and emitted a glimmering light to the computer science industry. Onward, with increasing generation of crowd sourced, globally contributed and engineered software, prominent platforms started to glow (Wordpress, Magento, etc), and from there on, with emergence of mobile computing, countless applications were produced for almost any purpose that involves in daily life. Together with web, and other digital gadgets, generating data, became a continual and smooth process. Where everyone could produce and release data about themselves, just by moving around passively, or by logging into PCs, mobiles, tablets, etc., an increasing amount of data became available to providers. According to IBM [28], a full 90% of all the data in the world has been generated over the last five years, and that confronts companies with a concept called ‘Big Data’. By the virtue of this concept, many researchers and practitioners started to emphasize on drawing patterns based on client generated data in different context, and that introduced a nascent factor, called ‘BDaaS’. Big data brought along prodigious frameworks, technologies and algorithms such as Apache Hadoop, Sparks, Flink, Storm, Sanza, MapReduce, and many more. The shape of business needs and expectations change dramatically, because many people have begun to realize what is possible.

Additionally, appliance of math for data at scale evolved into its new phase, and supplementary to that, prominent programming languages emerged to support better software engineering practices for parallel processing.

A. Global Shapeshift toward Data

In the shapeshift of the global industry, BDaaS is the core direction. Being the underlying essence for innovative and game changing business intelligence, BDaaS contrast with conventional data services architectures, by providing support for unstructured data, and vide variety of services, such as analysis and visualization. On December 2012, the world's authoritative market research institution Technavio, released "Global Big Data as a Service Market 2012-2016", it reported that companies need BDaaS solution to improve overall operational efficiency, innovative business strategies, IT system performances, employee performance, and the overall trajectory of the progression. Currently, many giant companies has engaged in BDaaS market space such as Microsoft, EMC, IBM, Amazon, Google, Snaplogic, SAP, Oracle, etc. Withall, BDaaS, is facing diverse challenges and limitations to face and overcome.

8. CHOKE POINTS

With all the hype, it is no wonder that organizations are getting into idea of having their own BDaaS initiatives. Nevertheless, as promising as the idea sounds, the reality is that over 70% of organizations fail to exert their big data goals [3]. BDaaS, which in another term is big data in the cloud, has surprised IT leaders with tremendous challenges. These choke points are taxonomized in three major groups, as following;

- Data Management and Analytics
- Security and Privacy
- Reliability
- Expertise Issues

Clearly, there is a disparity between these major areas of challenge, but the successful execution of big data on the cloud, requires a clear consideration of the aforementioned difficulties.

A. Data Management and Analytics

According to Ahmed and Saeed [27], challenges in the area of data management and analytics fall into two categories: capacity and performance. Cloud providers have to carefully consider the scaling capacity from a platform perspective. As data retention continues, the volume increase to double and triple year over year and that brings along major impact for capacity providers [29]. Cloud providers, has to obligate to be highly available, in order to scale in line with client's needs. In that obligation, there lies a vast amount of challenge to face. Capital expenditure required to respond to the scaling demand, is a not something any typical IT company can afford [30]. Moreover, operational challenges are the other factors that has to be considered. The foundational technology supporting every big data initiative is the Hadoop analytics platform. In addition to complexity of using the software, Hadoop regularly requires extensive internal resources to maintain. As a result, Hadoop utilizing companies wind up allocating the majority of their resources to the technology instead of the big data issues that are intended to solve. Recent surveys of data professionals, states that on-premise Hadoop is so intricate, that 73% of respondents felt that understanding the big data platform was the paramount challenge of a big data project [31]. Moreover, majority of organizations that adopt an on-premise Hadoop functionality fail to take into account that, sooner or later, their data storage and analytics demands are going to increase. This, in long-term, provides these companies with scaling issues. Hadoop analytics platforms rely on commodity servers, and that physical environment results in scalability problems and storage limitations. Hence, in order to overcome the challenge, more physical servers must be added, and that can be expensive, time consuming, and disruptive to the on-going project. Moreover, big data workloads are not on regular basis, which makes it a challenge to anticipate where resources should be located.

B. Security and Privacy

BDaaS remains in the hype circle of emerging technology. Nevertheless, despite the hysteria regarding the potential of big data, security and privacy challenges threaten to slow down the momentum [27]. By virtue of five V's of big data (velocity, veracity, variety, volume, and value) security issues and challenges has been amplified. These factors incorporate variables such as diversity of data sources, substantial cloud infrastructures, and formats, high magnitude of inter-cloud migrations, and streaming nature of data acquisition. As a result, traditional security mechanisms, which were tailored to secure small-scale static data often fall short.

The information security practitioners at the Cloud Alliance are aware of the fact, that BDaaS is the current trend.

By analysing the challenges of internal and external threats and review of current approaches to mitigate those risks, the Alliance's members presented with report detailing four major security and privacy challenges facing infrastructure providers and clients. These four major security and privacy challenges are as following;

- Existing encryption technologies that don't scale well to large datasets
- Real-time system monitoring techniques that works well on smaller volume of data but not very large datasets
- The growing number of devices, from smartphones to sensors, producing data for analysis
- General confusion surrounding the diverse legal and policy restrictions that lead to ad hoc approaches for ensuring security and privacy

Large data sets are distributed throughout the Big Data application; therefore, the entire data storage layer needs to be secured, and that puts forward a challenge for practitioners. To addresses these challenges, many mechanisms have been exerted such as Vormetric Encryption, Data Security Platform, Encryption, Key Management, Fine-Grained Access Controls, and Security Intelligence.

C. Reliability

In the distributed, complex, and inherently unreliable context of Cloud, approaches and mechanisms of increasing reliability of the Big Data are critical to the utilization of virtualized services within the practicing companies. In reality, Cloud providers need to be highly durable, decidedly available, and has to scale from a few bytes to exabytes. Today, the world's most reliable solution is Amazon's S3. According to service level agreement published by Amazon [32], S3 guarantees a 99.9% monthly availability and 99.999999999% durability per year. This is less than an hour outage per month. Thus, if, for example, customer stores 20,000 objects, it can be expected that two objects losses occur every 20,000,000 years on average. Amazon, being a giant role player, has achieved this reliability measure by storing data in multiple facilities with error checking and self-healing processes to detect and repair errors and device failures. Other Cloud providers may reach the same reliability measure too, but it requires an extensive capital expenditures and operational challenges. Global data cantered companies like Google and Facebook have the expertise and scale to implement similar infrastructures. Nevertheless, small and medium level companies, may not be able to expenditure the required capital, and therefore, fail in delivering reliable BDaaS. Besides the tremendous capital necessity, operational challenges are challenging too. BDaaS, in the Cloud, a naturally unreliable environment, undergoes various reliability challenges such as distributed data storage, inevitable occurrences of disk failures, and VM (virtual machine) failures. On the grounds of colossal architecture like Cloud, providing big data services reliably, underlying countless conditions and scenarios, is a very complicated matter. Discrete type of failures are probabilistic in the Cloud environment such as overflow failure, timeout failure, resource missing failure, network failure, hardware failure, software failure, and database failure. On the account of reliability analysis, Cloud is composed of various elements, and factors, that are each equally critical. These factors such as wide area networks, and heterogeneous software/hardware, data load and clustering, and scaling presents practitioners with considerable amount of complicated interactions among various sections of the Cloud. Ergo, current models are not matured enough to meet the market demands and to ensure the required reliability of this nexus of computer technologies.

D. Expertise Issues

Successful adoption of BDaaS is widely dependent upon getting the right people with the right skills. As BDaaS adoption accelerates, these people are getting harder and harder to find. BDaaS at minimum requires sophisticated teams of developers, data engineers, data scientists, quality engineers, and analysts who have the necessary knowledge and skills to identify actionable insights that create value and competitive advantage. Forming such a team, can be a challenging and expensive process, and failing to do so, results in major issues. The challenges to implement BDaaS are enormous, but so are the benefits. Overcoming those challenges, will guaranty a significant competitive advantage.

9. CONCLUSION AND RECOMMENDATION

The rise of the Internet provided a path for Big Data to ascent and grow. Gigantic volume of data is generated everyday around the world. This led to many issues and challenges regarding to varied nature of big data as well as handling this amount of data in business and industries. From the era of the mainframe to the era of the big data, our world has gone through distinct major phases of evolution. These phases has not only changed the approaches towards systems and data development and provisioning, but has also reshaped the overall conceptual framework toward resource utilization, value generation, and points of concentration. Evolutions in computer science brought along an increasingly service-oriented approaches toward system development and information provisioning, with emphasize on benefits of crowd’s unified wisdom, and mutual ingenuity. On that account, modern approaches laid the foundation of a prominent service-oriented platform, called BDaaS. BDaaS, being at hype of emerging technologies, derived from the analysis through Information Systems (IS), and Software Engineering (SE) theories and practices. Under the trajectory of BDaaS, business intelligent and analytics transmuted to catch up with leading edge technologies and to provide with increasing market demand.

However, with all the hype, major bottlenecks came as an obstacle to challenge various organizations and industries, practicing big data on the cloud. These bottlenecks surprised IT leaders with tremendous challenges in various areas such as data management, analytics, security, privacy, reliability, and expertise.

Taking all into consideration, it can be deduced that, BDaaS is still in infant stage and requires further researches and practices to parallelize itself with current market demands.

Based on that, a recommendation would be the, first of all, development of a reliable orchestration framework that scales well with the volume and velocity of the data, and second, a novel approach toward forging context driven knowledge models to materialize an actionable knowledge.

REFERENCES

[1] GRIT, "GreenBook Research Industry Trends (GRIT) Report," GreenBook2015

[2] M. A. Beyer and D. Laney, "The importance of'big data': a definition," ed: Gartner, 2012.

[3] J. Kaskade, "CIOs & big data: What your it team wants you to know," ed: Infochimps, 2013.

[4] M. Huberty, "Awaiting the second big data revolution: from digital noise to value creation," Journal of Industry, Competition and Trade, Vol. 15, pp. 35-47, 2015.

[5] J. Gantz, D. Reinsel, and C. Arend, "The digital universe in 2020: Big Data," Bigger Digital Shadows, and Biggest Growth in the Far East, IDC IVIEW, Sponsored by EMC Corporation, 2013.

[6] Z. Zheng, J. Zhu, and M. R. Lyu, "Service-generated big data and big data-as-a-service: an overview," in 2013 IEEE international congress on Big Data, 2013, pp. 403-410.

[7] C. Lam, Hadoop in action: Manning Publications Co., 2010.

[8] E. Xinhua, J. Han, Y. Wang, and L. Liu, "Big Data-as-a-Service: Definition and architecture," in Communication Technology (ICCT), 2013 15th IEEE International Conference on, 2013, pp. 738-742.

[9] J. Verble, "The NSA and Edward Snowden: surveillance in the 21st century," ACM SIGCAS Computers and Society, Vol. 44, pp. 14-20, 2014.

[10] H. Chen, R. H. Chiang, and V. C. Storey, "Business Intelligence and Analytics: From Big Data to Big Impact," MIS quarterly, Vol. 36, pp. 1165-1188, 2012.

[11] S. Kudyba and M. Kwatinetz, "Introduction to the Big Data Era," Big Data, Mining, and Analytics: Components of Strategic Decision Making, p. 1, 2014.

[12] F.-Z. Benjelloun, A. A. Lahcen, and S. Belfkih, "An overview of big data opportunities, applications and tools," in Intelligent Systems and Computer Vision (ISCV), 2015, 2015, pp. 1-6.

[13] D. Howe, M. Costanzo, P. Fey, T. Gojobori, L. Hannick, W. Hide, et. al., "Big data: The future of biocuration," Nature, Vol. 455, pp. 47-50, 2008.

[14] R. Akerkar, "Big Data & Tourism," 2012.

[15] S. Yousfi and D. Chiadmi, "Big Data-as-a-service solution for building graph social networks," in Cloud Technologies and Applications (CloudTech), 2015 International Conference on, 2015, pp. 1-6.

[16] P. Malik, "Governing big data: principles and practices," IBM Journal of Research and Development, Vol. 57, pp. 1: 1-1: 13, 2013.

[17] N. Mahmood, Cloud computing: Springer, 2013.

[18] E. S. Group, "IG DATA-AS-A-SERVICE: A Market and Technology Perspective " MC Corporation 2012.

[19] C. Prokopp. (2014, The four types of Big Data as a Service (BDaaS). Big Data Science and Cloud Computing

[20] J. Buckley, "How to Choose a Big Data-as-a-Service Company," ed, 2015.

[21] P. Mell and T. Grance, "The NIST definition of cloud computing," 2011.

[22] R. Kazman and H.-M. Chen, "The metropolis model a new logic for development of crowdsourced systems," Communications of the ACM, Vol. 52, pp. 76-84, 2009.

[23] X. Xu, Q. Z. Sheng, L.-J. Zhang, Y. Fan, and S. Dustdar, "From big data to big service," Computer, Vol. 48, pp. 80-83, 2015.

[24] Bloomberg Businessweek Research Services, "The Current State of Business Analytics: Where Do We Go From Here?," Bloomberg and SAS2011.

[25] D. D. Rustam and T. P. T. van der Weide, "Towards a methodological framework for designing a knowledge market," in Advanced Informatics: Concept, Theory and Application (ICAICTA), 2014 International Conference of, 2014, pp. 232-237.

[26] D. Ju and B. Shen, "On building knowledge cloud," in Computer Science and Service System (CSSS), 2011 International Conference on, 2011, pp. 2351-2353.

[27] E. S. A. Ahmed and R. A. Saeed, "A Survey of Big Data Cloud Computing Security," International Journal of Computer Science and Software Engineering (IJCSSE), Vol. 3, pp. 78-85, 2014.

[28] IBM. (2016). Bringing big data to the enterprise. Available: https://www-01.ibm.com/software/data/bigdata/what-is-big-data.html.

[29] I. D. Constantiou and J. Kallinikos, "New games, new rules: big data and the changing context of strategy," Journal of Information Technology, Vol. 30, pp. 44-57, 2015.

[30] A. Depeige and D. Doyencourt, "Actionable Knowledge As A Service (AKAAS): Leveraging big data analytics in cloud computing environments," Journal of Big Data, Vol. 2, p. 1, 2015.

[31] J. King and R. Magoulas, 2015 Data Science Salary Survey: Tools, Trends, What Pays (and What Doesn't) for Data Professionals, Third ed. United States of America: O'Reilly Media, 2015.

[32] Amazon. (2016, 24 November). Service Level Agreement - Amazon Simple Storage Service (S3). Available: https://aws.amazon.com/s3/sla.