Management of geo-distributed intelligence: Deep Insight as a Service (DINSaaS) on Forged Cloud Platforms (FCP)

Kaya Kuru*,[a]

[a]School of Engineering, University of Central Lancashire, Fylde Rd, Preston, UK, PR1 2HE

Abstract

The recent advances in the cyber-physical domains, cloud and edge platforms along with the advanced communication technologies play a crucial role in connecting the globe more than ever, which is creating large volumes of data at astonishing rates. Data analytic tools are evolving rapidly to harvest these explosive increasing data volumes. Deriving meaningful insights from voluminous geo-distributed data of all kinds as a strategic asset is fuelling the innovation, facilitating e-commerce and revolutionizing the industry and businesses in the transition from digital to the intelligent way of doing business with globally generated distributed intelligence. In this perspective, in this study, a philosophical industrial and technological direction involving Deep Insight-as-a-Service (DINSaaS) on Forged Cloud Platforms (FCP) along with Advanced Insight Analytics (AIA), primarily motivated by the global benefit is systematically analyzed within sophisticated theoretical knowledge, and consequently, a geo-distributed architectural framework is proposed to 1) guide the national/international leading organizations, governments, cloud service providers and leading companies in order to establish an environment in which exponentially increasing voluminous big data can be harvested effectively and efficiently, 2) inspire the transformation of big data into wiser abstract formats in Specialized Insight Domains (SIDs) in order to help make better decision making and near-real-time predictions, especially for applications requiring low-latency, and 3) direct all the stakeholders to rivet the high-quality products and services within Automation of Everything (AoE) by exploiting continuously created and updated insights in dedicated specialized domains within geo-distributed datacenters located in geo-distributed cloud platforms.

Key words: Cyber-physical domains (CPD), Automation of Everything (AoE), Forged Cloud Platforms (FCP), Deep Insight as a Service (DINSaaS), Advanced Insight Analytics (AIA), Specialized Insight Domains (SIDs), Sanitization

1. Introduction

Recent advances in cyber-physical domains and platforms in which physical objects around us become an irrevocable part of the global information system and Big Data (BD) have created the chance to develop an open architecture with effective sharing and intelligent services [1]. Today, the world is connected more than ever and the "Business Intelligence (BI)" landscape enabling improved and optimized decisions and performance by leveraging analytics software to transform data into intelligence [2] is dominated by intelligent inferences acquired from BD using Business Analytics (BA) that is the practice and art of bringing quantitative data to bear on decision-making [3]. The next level of business analytics, now termed BI, refers to data visualization and reporting for understanding "what happened and what is happening" [3] and most importantly "what is going to happen". Online retailers are rapidly adopting BD analytics solutions, particularly predictive analytics aiming to predict the market and customer needs [4]. The era of BD has attracted much attention and accelerated the use of BD analytics and new advanced analytical tools and techniques are on the rise to support innovation, promote productivity, improve efficiency, and manage the intelligent autonomous traffic on the global network, cloud platforms, and smart domains. For some companies, like Facebook, Twitter, and LinkedIn, nearly the entire value of the company lies in the analytic and predictive value of the voluminous data from their social networks; Facebook was worth more than double General Motors and Ford combined even though they manufacture no products and sell no services in the traditional sense to their users [3]. Amazon and Pandora, use so-

*Corresponding author

Email address: kkuru@uclan.ac.uk (Kaya Kuru)

cial network data as important components of predictive engines aimed at selling products and services [3] to targeted customers. Change for the better requires change for the way of doing business intelligently. Wisdom and knowledge extraction and information harvesting on distributed platforms are the main developing and promising subjects of doing business intelligently in today's business environment. However, there is chaos in the effective and efficient management of the globally distributed knowledge and wisdom regarding various cloud platforms, numerous smart domains, privacy and security concerns, and insufficient rules and regulations. As a result, it is highly unlikely to create a synergistic environment in which the globally created voluminous data and intelligence can be exploited effectively and efficiently. Building up a global synergistic intelligent infrastructure requires orchestration of resources in a new concept that is able to mitigate the chaos by generating, disseminating and acquiring desired insights from globally generated voluminous data within Automation of Everything (AoE) using Advanced Insight Analytics (AIA). AoE was analyzed based on the cutting edge technologies, Advanced Mechatronics Systems (AMSs) and consequently, a framework of AoE was proposed for the first time in the study [4]. In this framework, all the stakeholders and advanced devices can be connected to each other in the Internet of Everything (IoE) environment intelligently via autonomous machine-to-machine (M2M), Peer-to-Peer (P2P) communication to be able to work as a part of bigger systems in which bigger synergies regarding location independent-monitoring, control and actuation are generated in order to both build cutting edge technologies and gain competitive advantages in the market.

The data stored on the cloud platform might be compromised. The studies conducted by ABCNews and Boston Globe show that it is achievable to infer the sexual orientation of a user through mining a Facebook subnetwork involving the user's friendship relations, gender, and other attributes [5]. There are many studies related to data security and privacy issues on the cloud platform to alleviate these similar concerns. Encryption was found to be the most widely applied technique [6] for protecting highly sensitive data such as passwords, physical locations, sexual orientation, names, ID numbers, images, personal files, bank transactions from unauthorized access by all entities, including service providers, which in turn makes the third parties not able to reach and analyze most of the data on the cloud platforms. New effective approaches are needed to be established to open this BD to everybody in order to unveil its potential and economic value without compromising the privacy and security concerns.

We are overwhelmed with data [7]. Large volumes of BD being generated exponentially in different formats are in the geo-distributed cloud platforms and likely input for all other smart systems and enterprises as insights, which will contribute to the smooth working of these systems and enterprises substantially. As the amount of BD being processed on datacenters in multiple cloud platforms increases, the network resource consumption also increases and BD management across multiple datacenters in multiple cloud platforms is an important and challenging task [8]. Streaming of this exponentially increasing voluminous data at once may choke the underlying current network infrastructure [8]. Recognizing the growing demand for ways to handle geo-distributed data cost-effectively, researchers have begun to focus on how to efficiently analyze geo-distributed datasets [9]. However, many solutions address only how to store data across datacenters and few efforts have investigated how to effectively compute with it [9]. In cloud platforms, 1) most of the time BD can not be reached because of the privacy and security issues and effective tools are being deployed to detect and respond faster to cyber threats, attacks, breaches of data, 2) This BD should be before published in the public domains to be explored and exploited for further analysis and purposes— not to mention that data still carry high risks of leaking sensitive information, and moreover, 3) processing a substantial amount of data within a very small time interval is a great challenge for low-latency cloud applications [10] where analysis of BD through the geo-distributed datacenters incur huge communication cost, particularly where BD is needed to be collected from geo-distributed datacenters and stored locally to be processed using the current processing techniques such as Apache Hadoop. Therefore, new approaches are needed first to reduce privacy and security risks to minimum, and second to make the most out of BD regarding extracting thorough and up-to-date insights in a timely manner. In this perspective, a new approach is designed in this study to cover those concerns in a holistic concept.

There is no study in ScienceDirect repository directly related to Insight-as-a-Service (INSaaS), Wisdom-as-a-Service (WaaS) and we have only found two articles in IEEE repository about WaaS, which is quite surprising. The titles of these articles are "WaaS: Wisdom as a Service" [1] and "Wisdom as a Service for Mental Health Care" [11]. What makes the issue worse is that there is no study in the literature that analyzes and utilizes extracted insights within Automation of Everything (AoE) involving all the cloud platforms and smart domains.

We have noticed that this important subject has not been well covered by scientists and researchers regarding the common demand from everyone in the context of "we are drowning in data, but starving for knowledge"; "no quality public domain to establish a quality decision-making platform for a specific field". The importance of this research is derived from the fact that exploiting insights within AoE is going to be immensely focused in the following years regarding the development of advanced communication technologies (e.g., 5G), cyber-physical systems, smart platforms and domains [4]. In this regard, to fill this gap, a philosophical industrial and technological direction is analyzed rather than focusing on technological details, and a framework is proposed in order to guide the national and international leading organizations, governments and leading companies, particularly, IT-based companies to play a crucial role in strategizing intelligent way of doing business. To clarify the novelty of this paper, the contributions are outlined as follows.

1) This paper focuses on how to make BD accessible to everybody by alleviating the privacy and security concerns in order to generate the likely insights to be explored and exploited by any entity requiring data-driven decisions and actuation.

2) For the first time, a new term, Forged Cloud Platforms (FCP) is highlighted to indicate the virtual integration of all the geo-distributed cloud platforms, smart domains, BD created by the leading companies, distributed and centralized computing and storage units in a new concept in which Very Very Very BD (VVVBD) is scaled and globally created insights can be aggregated in specialized wise domains under a unique virtual platform with accessible services enabling timely advanced insights.

3) For the first time, another new term, Deep Insight as a Service (DINSaaS), is used to refer to the insights harvested from globally created insights aggregated and specialized in dedicated wise domains, and the essential features of DINSaaS along with an explicit definition is revealed within a framework proposed in regard to the newly proposed global smart data management infrastructure, FCP.

4) For the first time, another new term, Advanced Insight Analytics (AIA) along with Specialized Insight Domains (SIDs) is used to indicate a new way of gaining further insights, so-called DINSaaS by exploring and exploiting the continuously created and updated insights along with the preprocessed raw data in the dedicated specialized wise domains, so-called SIDs on FCP with effective privacy and security preserving abilities.

5) To the best of my knowledge, this is the first attempt that explicitly studies DINSaaS and gives concrete recommendations about how to gain most valuable insights and put them in the hands of users by forging the features of the cyber-physical systems, cloud platforms, smart domains, communication technologies, particularly promising wireless communication technologies using intelligent M2M communication, and Industry 4.0 (4IR) in a new concept - Automation of Everything (AoE) on the infrastructure of FCP, to enable the implementation of next-generation autonomous smarter systems.

6) The frameworks for FCP, SIDs, DINSaaS and effective use of AIA are identified and new guidelines are unfolded to increase the efficacy of these new concepts as voluminous data piles up, which may have a great impact on the economy by transforming enterprises and devices into a new perspective in which the global wealth can be increased substantially benefiting the whole globe.

The remainder of this paper is organized as follows. The background is analyzed in Section 2. Section 3 introduces the preprocessing and organization of BD and VBD in VVBD pools for FCP along with the essential features of SIDs. Section 4 describes FCP whereas DINSaaS along with AIA is explored with respect to FCP in Section 5. The challenges are unfolded in Sections 6. The discussion is provided in Section 7. Finally, Section 8 draws conclusions and provides directions for potential future ideas.

2. Background

2.1. Cloud, edge/fog and Cyber-Physical Systems (CPS)

First, I would like to cover the main technological concepts briefly related to this study such as cloud platforms, edge/fog platforms and Cyber-Physical Systems (CPS) [1] The cloud platform with vertically expandable data storage and processing capabilities has the advantages for massive storage, heavy-duty computation, global coordination, and wide-area connectivity, while edge/fog, particularly mobile-edge computing (MEC)[12] is useful for real-time processing, rapid innovation, user-centric service and edge resource pooling [13]. Each cloud platform is being expanded with main distributed advanced datacenters around the globe where each one contains tens of thousands, in some

[1] Interested readers are referred to the study [4] for more detailed information about smart platforms and CPS.

cases, hundreds of thousands of servers called supercomputers using High-performance computing (HPC) systems and tens of thousands of high-bandwidth endpoints. The cloud is approaching the edge as the massive network in a wider infrastructure along with the deployment of multiple virtual machines (VMs) through virtualization is being constructed by the leading providers [14], in particular, using the smaller versions of cloud platforms, i.e., cloudlets, which enables reduced latency. By leveraging low-latency offload, cloudlets enable a new class of real-time cognitive assistive applications on mobile cloud convergence [15] by massive virtualization with VMs managed by using hypervisors servers [16]. The main cloud service providers are IBM, Amazon EC2, Microsoft Azure, Fiware and Google providing an open public network connecting businesses, individuals, organizations, and governments all around the world under an umbrella with Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS) using the pay-per-use model. The start-up costs of enterprises, particularly emerging ones, on the cloud are quite small with respect to starting a rapid worldwide business from scratch, which makes this platform very appealing along with the advanced abilities of 1) automatic detection of compromised accounts, malware, data breaches, and malicious insiders, and 2) reducing the risk of data loss using redundancy, archiving, VM backups at multiple locations.

Cloud networking uses 1) the Software-Defined Networking (SDN) [17] enabling the configuration of the datacenter at a high level with much less human intervention and allocation of VMs, virtual networking, and virtual storage to a new tenant with specified service level guarantees, and 2) specialized cloud operating systems (e.g., OpenFlow, OpenStack) to manage server, storage, and networking resources in order to supports multiple applications from third parties. Interested readers are referred to the study [18] for topological connections and physical component categorizations in cloud computing, to the studies [17], [19] for cloud and network orchestration in SDN datacenters. Edge (or fog) or most recent popular platform, so-called MEC is an emergent architecture for computing, storage, control, and networking that distributes these services closer to end-users [13] to enable a more independent processing and organization, particularly for applications requiring real-time decision making, low-latency, ultra-low-latency, high privacy, and security. The segmentation of what tasks go to the fog/edge and what tasks go to the back-end cloud is application-specific and can change dynamically based upon the state of the network, including processor loads, link bandwidths, storage capacities, fault events, and security threats [20].

Internet of Things (IoT) with resource constraint characteristics is composed of physical objects embedded with electronics, software, and sensors, which allows objects to be sensed and controlled remotely across the existing network infrastructure, facilitates direct integration between the physical world and computer communication networks, and significantly contributes to enhanced efficiency, accuracy, and economic benefits [21]. With IoT, physical objects are seamlessly integrated globally so that the physical objects can interact with each other and to cyber-agents in order to achieve mission-critical objectives [22]. IoT envisions a future in which digital and physical entities can be linked, by means of appropriate information and communication technologies. It's predicted that by the year 2020 about 75 billion devices would be connected around the world [23]. The number of connected things will exceed 7 trillion by 2025 which makes 1000 devices per person and an estimated value of 36 trillion of dollars [24]. The emergence of IoT has led to increasing data volumes [25], which is expected to grow exponentially to 44 zettabytes by 2020 [26] accounting for 10% of the total digital universe [27]. One of the primary problems in cloud computing today is how to manage sensitive workloads running on the cloud [28]. Service providers are trying to assure their customers about this issue using various approaches, techniques, and policies involving their users with Shared Responsibility Model (SRM) generally promoted by Amazon Web Services (AWS) and Elastic Compute Cloud (EC2) [2] on the cloud and edge/fog platforms.

2.2. VBD created by smart domains, BD created by leading companies and organizations, and BD analytics

The world's internet population is growing significantly year by year; as of January 2019, the internet reaches 56.1% of the world's population representing 4.39 billion people - a 9% increase from January 2018 [30]. IDC (a technology research firm) estimates that data has been constantly growing at a 50 percent increase each year, more than doubling every two years [1]. With an increased digital consumption the world is creating massive amounts of data on a daily basis [31]. According to Domo's Data Never Sleeps 6.0 report, there are 2.5 quintillion bytes (1 million terabytes) of data created each day and more than half the

[2] Readers can find more information about the interplay between the IoT, cloud and edge/fog in [13], [29].

world's web traffic comes from smartphones, and it's predicted that 6.1 billion people will have access to a smartphone by 2020 [31]. We can safely conclude that more than 90% of the existing BD in the world has been generated in the last several years. The main features of BD, 5 V's, are 1) Velocity (i.e., speed at which tremendous amounts of data are being generated, collected and analyzed), 2) Volume (i.e., tremendous amounts of data are being generated each minute), 3) Value (i.e., worth of the data), 4) Variety (i.e., different types of data in different formats), and 5) Veracity (i.e., quality and trustworthiness of the data). Most of the BD is unstructured (e.g., text, speech, video) around 85-90% regarding the variety, which makes the analysis and interpretation so difficult in gaining insights even though there are promising attempts to develop new tools (e.g., text mining, web mining, image mining, social network analysis (SNA)) that are able to analyze unstructured BD.

The concept of BD has fundamentally changed the way organizations manage, analyze and leverage data in any industry [32]. Bigger data means more insights and better decision making. For instance, big healthcare data has considerable potential to improve patient outcomes, predict outbreaks of epidemics, gain valuable insights, avoid preventable diseases, reduce the cost of healthcare delivery and improve the quality of life in general [32].

In this section and in the rest of the paper, in order to visualize the proposed concepts explicitly, BD refers to the data created by individual leading companies whereas Very BD (VBD) corresponds to the data in the smart domains (e.g., data in the smart city). Very Very BD (VVBD) is used for the data in the individual cloud platforms (e.g., Google Cloud) whereas Very Very Very BD (VVVBD) corresponds to the combined data of all cloud platforms.

2.2.1. VBD created by smart domains

Some of the main smart domains are smart city, smart home, smart building, smart transportation, smart health, smart industry, smart factory, smart shopping and manufacturing, smart logistics and retail, smart energy and smart grid, and smart agriculture. The connected IoT devices or mechatronics devices in these smart domains not only talk to each other within their smart domains, but also they can talk to other devices in the other smart domains as well, e.g., security, fire or gas alarm using intelligent sensors in the smart home domain may trigger an action for police or fire department in the smart city domain [4]. More explicitly, there are no strict boundaries between these smart domains; some of the outputs of the smart devices may input for other

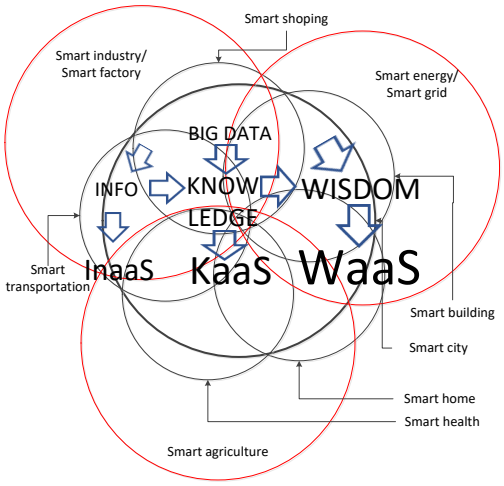

Figure 1: Interaction of smart domains and formation of WaaS.

smart devices within both their domains and other domains, as illustrated in Fig 1, in which the smart city is in the center to indicate people-focused cyber-physical understanding, since more than 60% of the population will be living in urban environment by 2030 [33] and the global population is expected to double by 2050 [34]. Cities with heavy populations escalate burden on energy, water, buildings, public places, transportation and many other things [33]. The proliferation of devices with communicating-actuating capabilities is bringing closer the vision of an IoT, where the sensing and actuation functions seamlessly blend into the background and new capabilities are made possible through the access of rich new information sources [35]. Interested readers are referred to the study [4] for more detailed information about the smart domains with many examples.

A voluminous data that we name it as VBD, is created in these domains. Still, there are serious restrictions in sharing of data between these domains regarding privacy and security concerns where data can be reached from anywhere anytime with worldwide distributed computing environments, and sharing of data between these domains within an effective and efficient infrastructure addressing cyberattacks concerns is required, which would make life functional and easier in many aspects. In this regard, one of the recent prominent trends is to integrate all smart domains in a combined architecture of the cloud platform [36] and a revolutionary networking model called Information-Centric Networking (ICN) has recently attracted the attention of the research community working on data dissemination in various smart domains [37]. A uni-

fied framework — Smart and Connected Communities (SCC) was presented for better preservation and revitalization (the needs for remembering the past), liveability (the needs of living in the present), and sustainability (the needs of planning for the future) in the study [34] aiming to integrate big cities, small towns, and beyond in order to establish better working systems on better decision-making abilities using BD. Similar data-sharing attempts between the smart domains will increase in the following years where privacy and security concerns are addressed well, particularly using INSaaS as explained in Section 3.

2.2.2. *BD created by leading companies and organizations*

Facebook has 1.65 billion users with 1 billion active users per month, Twitter has 600 million users with 0.5 billion tweets published per day, Amazon has 304 million users with 9.65 billion items traded per year, Tencent QQ has 829 million active users with up to 210 million simultaneous online users, WeChat has over a billion users with 700 million active users [5]. BD created every minute by leading companies and everybody across several industries, including tech, media, retail, financial services, travel, and social media is presented in Fig. 2. With such a large scale and variety of data, Social Network Analysis (SNA) becomes increasingly crucial for classifying end users, predicting buying interests, foretelling event occurrence [5]. As in smart domains, the data sharing between these companies is highly limited regarding privacy, security issues, insufficient rules and regulations, and competitive commercial concerns, which extremely reduce the chance of unfolding and gaining numerous insights, and consequently a tremendous amount of economic value is unfortunately lost.

2.2.3. *BD Analytics*

Analytics is the science of using data to build models that lead to better decisions that in turn add value to individuals, companies, and institutions [38]. More and more sensors and devices are being interconnected via IoT techniques, and these sensors and devices will generate massive data and demand further processing, providing intelligence to both service providers and users [39]. Exponential increase in the volume of data being created and analyzed has triggered interest in a new interdisciplinary form of scientific inquiry referred to as "data science" and "data analytics"[40] which has given rise to a new profession: the data scientist with analytics skills [3] in order to make the tremendous potential explicit, i.e., secrecy of life, in BD and exploit

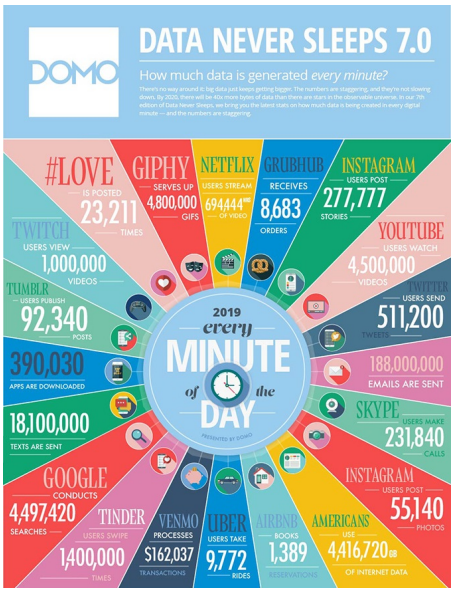

Figure 2: Data generated every minute (Courtesy of DOMO).

it as a profitable business. Data science is a mix of skills in the areas of statistics, machine learning (ML), math, programming, business, and IT [3]. BD created in the cloud platform as a Data-as-a-Service (DaaS) turns into Information-as-a-Service (InaaS); InaaS turns into Knowledge-as-a-Service (KaaS) using BD analytics, and finally, KaaS transforms into Wisdom-as-a-Service (WaaS) where Wisdom is insight to know what's true or right for making correct judgment, decisions, and actions. [1]. To summarize, BD is turned into IaaS and WaaS to be employed to gain better insights. WaaS is created both in edge and cloud platforms. The WaaS standard and service platform is expected to be fine-tuned continuously as a core infrastructure for intelligence industry and smart city to support the development of various intelligent IT applications and it is anticipated that this will bring a huge economic value for intelligence IT industry by realizing the pay-as-you-go concept [41].

More than 87% of organizations are classified as having low BI and analytics maturity, according to a survey [42]. Another survey conducted by MIT Sloan Management Review found that organizations that "strongly agreed that the use of business information and analytics differentiates them within their industry" were more likely to be top performers and such organizations use analytics in a wide spectrum of decision making, both to guide future strategies and for day-to-day operations [2]. The digital business future confronts individuals and companies with almost unlim-

ited possibilities to create business value through data and analytics by deploying and running BD analytics on the cloud [43], over 60% of global enterprises will adopt public clouds for BD analytics by 2020 [44]. We can safely conclude that the way of doing business regarding CPS triggers a new industrial revolution, mainly based on the insights that can be gained from the globally generated BD from multiple channels using BD analytics where bigger data increases the chance of representing the real world better and consequently extracting better insights for better decision making. The BD analytics with its decision support capability, provide critical information such as historical reports, statistical analyzes, time-series comparison, forecasting business opportunities and executive summaries to managers and executives to facilitate better decision making [45]. Furthermore, the BD analytics produce and deliver insights to be used as an input into other systems or merged with other insights created locally for better working IoTs and AMSs to make autonomous and dynamic real-time decisions. The BD analytics are capable of collecting the scattered data to understand user behavior and preferences from multiple perspectives to portray an integrated picture [46]. Moreover, subscribers' living habits and the timetable can be generally inferred from the usage of traffic over different time periods of a day; their surfing habits and interests can be roughly obtained from the logs; their frequently visited places or the range of activities can be approximately derived from home location register (HLR) databases [46]. Within this concept, intelligent systems such as smart home applications mimicking the users' behaviors can be established to create more comfortable life with customization of environment according to daily changing habits and activities.

Most recent data analytic tools designed to work on the cloud platforms are analyzed in [47]. Cloud analytics enables businesses to carry out analytics through an integration of hosted data warehouses, BI, and other analytics [2]. There are several BD analytics infrastructures along with their software provided by leading cloud service providers to enable processing large volume of data such as Big Query as Database-as-a-Service (DaaS) provided by Google, No-SQL, BigInsights provided by IBM, Apache Hadoop, Apache Spark enabling interactive SQL on Hadoop, Hive built based on the Hadoop Distributed File System (HDFS) with a master-slave (i.e, JobTracker-TaskTrackers) architecture enabling ad hoc query processing, Elastic MapReduce provided by Amazon. The components of the most commonly used BD analytics, Hadoop in a broader perspective are 1) high-level languages (i.e., Pig (execution framework), Cascading, Hive (data warehouse)), 2) execution engine (i.e., MapReduce), 3) distributed light-weight database (i.e., HBase), 4) distributed file system (i.e., HDFS), 5) centralized tool for coordination (i.e., Zookeeper). Pig is used by Yahoo; Hive with SQL-like language called HiveQL is used by Facebook; and Jagl with the ability of processing structured and nontraditional data is used by IBM. Hadoop has no support for geo-distributed data processing [9] even though it has a Java-based software framework for distributed processing of large datasets across large clusters of computers with limitless concurrent tasks. In contrast to Hadoop's two-stage disk-based MapReduce paradigm, Spark's multi-stage in-memory primitives provides performance up to 100 times faster for certain applications [9]. Within Apache Hadoop (Hadoop-as-a-Service), BD is collected from geo-distributed datacenters and locally stored in HDFS to be processed by MapReduce software computing, which makes it highly difficult to gain timely insights, more importantly it gets more difficult to store BD locally as the data volume grows exponentially regarding the restricted computing abilities, hardware limitations such as processing power, network bandwidth, storage limitations, and high-latency. The bandwidth availability between different datacenters significantly varies over time which usually is the bottleneck of such an evaluation [48], particularly for low-latency requirements. Raw BD stored in no-sql database are messily scattered and can't be used directly for two reasons [34]: first of all, for most of them, data cannot be interpreted by the model itself, and additional features have to be handled in the application logic [34]. Secondly, the overwhelming majority of BD are few of value while only a drop in the bucket is valuable [49].

Some of the other BD analytics projects established to solve the different problems are as follows: Ambari (cluster management), Avro (data serialization), Cassandra (multi-master database), Chukwa (data collection), Hbase (distributed database), MaHout (ML and data mining (DM)), Tez (data-flow programming framework intended to replace MapReduce), Cloudera Hortonworks, Microsoft (HDInsight), MapR, Map Altiscale, Factor in Apache storm (stream processing) and Kafta. Data analytics can be used on fog/edge platforms for processing large volume of multi-modal and heterogeneous data from various sensor devices and other IoT devices to achieve real-time and fast processing for decision making [50] where the processing power with increasing storage units is getting bigger locally. In other words, local advanced Hybrid Cloud-Edge Analytics (HCEA) should enable performing local analyt-

ics, identifying usable information from raw data, extracting insights in abstract forms and finally transmitting the result to the cloud platform.

2.3. Sanitization of BD and cybersecurity risks

Large amounts of data stored on the cloud are very sensitive, and so data privacy remains one of the top concerns for many reasons; mainly those relating to legal or competition issues [51]. In a Gallup poll, 27% of respondents said they or someone within their household had credit card information stolen [52]. 1 billion user accounts in Yahoo were compromised in 2013 [53]; Hackers attacked on Apple's iCloud platform that resulted in the release of the personal photographs of many high profile figures in 2014 [54]. LinkedIn, Sony, Oracle, T-Mobile, Dropbox and many others were also attacked similarly by hackers. By 2020, more than 70% of enterprises will continuously monitor for sensitive data incidents [43]. In this sense, the BD analytics technologies are instrumental for organizations to improve their capabilities in discovering potential threats, detecting actual threats, gathering intelligence about attacks, and deploying a comprehensive response to minimize the business impacts of cyberattacks [55] such as theft of credit card data, trade secrets. In this perspective, privacy-preserving data analysis is an emerging discipline within data science, which posts several challenges currently being simultaneously tackled from several areas such as hardware, systems security, cryptography, statistics, and ML [56]. Several privacy-enhancing techniques evolved in the last decade have different trade-offs, maturity levels, and privacy guarantees, and in some cases solve slightly different problems [56]. By 2020, large global enterprise use of data masking or similar pseudonymization techniques will increase to 40%, from 10% in 2016 [43].

Sanitization, e.g., anonymization/de-identification, and cybersecurity measures to prevent breaches of sensitive information allow the sharing of data for secondary purposes, such as research, the establishment of decision-making tools, extraction of other meaningful information that can be an input to other systems. Sharing of data should protect individual privacy, but still ensures that the data is of sufficient quality that the analytics are useful and meaningful [57]. In this manner, new sanitization approaches are on the rise to protect the privacy and security in addition to conventional most commonly used sanitization techniques such as k-anonymity [58], privacy [59] and l-diversity [60]. Various new sanitization approaches specific to BD on the cloud and edge platforms have recently been extensively introduced in order to both mitigate the shortcomings of existing ones and process specific types of BD effectively and efficiently such as 1) LinkMirage to address the link privacy in the social media data [61], 2) HCMPSO in an IoT Environment [62], 3) data-sanitization for preventing sensitive information in social networks involving various data-manipulating methods [63], 4) automatic unsupervised general-purpose sanitization of textual documents by detecting and hiding sensitive textual information while preserving its meaning [64], 5) collaborative search log sanitization toward differential privacy and boosted utility [65], 6) ant colony system sanitization approach to hiding sensitive item sets — ACS2DT, in order to hide sensitive and critical information by decreasing sanitization side effects and enhancing the performance of the sanitization process [66], and 7) individual trajectory data sanitization — Lclean, using a plausible replacement method [67].

A study by Skyhigh Networks, a cybersecurity firm, found that 18.1% of all documents uploaded to cloud-linked systems contain sensitive data [68]. Zhang *et al.* demonstrated that 20% of the big image data was found sensitive and maintained on the edge platform whereas 80% was found non-sensitive and encrypted, then, subsampled and stored in the cloud platform [69]. Another research estimates that 90% of the data generated by the endpoints will be stored and processed locally rather than processed in the cloud [70] where sending all the data to the cloud requires prohibitively high network bandwidth [13]. New studies that aim to sanitize data at its source before sending to cloud platforms are emerging such as [26] in which a privacy-preserving smart home system, which connects a single home controller with data-hiding capabilities through community networking and integrates the data to a hierarchical architecture on a cloud platform for a data analytical access control mechanism. In similar ways, sanitization starts on the edge platform. However, quite an important amount of sensitive data is placed in the cloud platforms even though most of the sensitive data can be processed and maintained on the edge platforms. In this regard, sensitive data, in particular, private data on the cloud platform such as sensitive personal data, medical data, credit card information and transactions should be managed carefully regarding the privacy and security aspects, particularly cyberattacks. What happens if a smartphone operating as an edge platform is hacked by a cyber attacker; cameras that are meant for surveillance may turn into cameras that are violating our privacy [4]. There have been various cases in which the personal identities of the owners of the sensitive data have been unveiled on the datasets placed in the pub-

lic domain for several reasons such as research [7]. For instance, when the state of Massachusetts released medical records summarizing every state employee's hospital record in the mid-1990s, the Governor gave a public assurance that it had been anonymized by removing all identifying information such as name, address, and social security number and he was surprised to receive his own health records (which included diagnoses and prescriptions) in the mail [7]. 50% of the Americans can be identified from city, birth date, and sex; 85% can be identified if you include the zip code as well [7]. You will probably be left with nothing useful if you really do remove all possible identification information from a database [7]. Sanitization of the data is not an easy process and it requires data scientists expertized particularly in sanitization to address the privacy and security concerns in order to mitigate the possible risks. Security is "confidentiality, integrity and availability" of data whereas privacy is the appropriate use of user's information [32]. Anonymization needs to be more than simply masking or generalizing certain fields-anonymized datasets need to be carefully analyzed to determine whether they are vulnerable to attacks [32]. In addition to sanitization, laws, and regulations should be amended to ensure the privacy and security wherever breaches emerge, which is not within the scope of this study and not explored in this paper in detail.

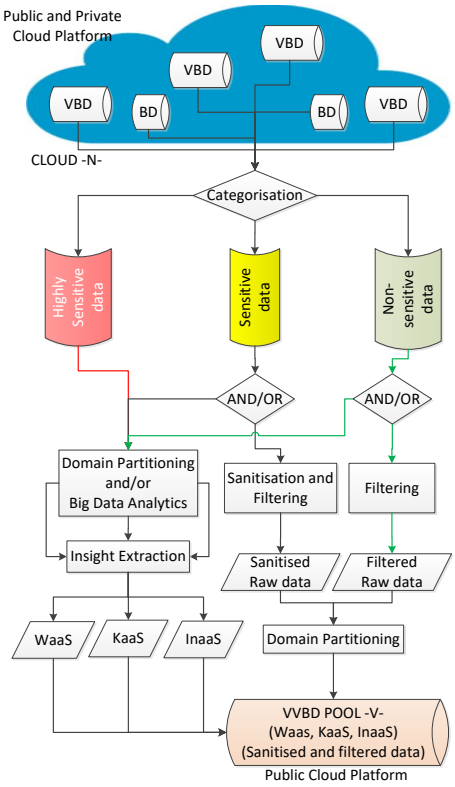

Figure 3: Categorization and processing of VBD for VVBD pool.

3. Preprocessing and organization of BD and VBD in VVBD Pools for FCP

3.1. Transforming of BD and VBD into raw data and abstract forms in VVBD Pools

The threat of predicting sensitive information become now a serious issue [5] even after the sanitization process. Owners of data are concerned with the risks of unauthorized usage of their sensitive data by various entities, including service providers [71] on the cloud platforms, particularly on the private cloud platform. Therefore, they avoid sharing their data with the outside community. Processing of highly sensitive data or sensitive data should be consent-based to be used for secondary purposes. In order to mitigate the concerns, the cloud service providers should cooperate with the owners of data, particularly, owners of highly sensitive data to be able to process their data and carry out sanitization operations to extract insights and to publish these insights along with sanitized data in public domains. To do this, first, owners of data should be convinced that their privacy and security will not be breached with the usage of strong advanced analytical tools, and second, they should be paid for this process as an incentive, which will help take their consent and increase the amount of shared data in public domains. National and international rules and regulations should be tailored to protect the owners of data. Extraction of insights, in other words, WaaS, KaaS, and InaaS should be carried out by the service providers to avoid any cyberattacks with the consent of the owners of the data; or alternatively, analytics under the control of the organizations with an ability to work on encrypted data using a number of parameters can be employed by the researchers as a gate to reach and analyze BD, which improves the effectiveness of cybersecurity capabilities without violating the privacy of the individuals. Encrypting data incurs additional storage and query processing costs — computationally expensive. Additionally, cryptographic schemes and practical systems analyzed in a recent study by Moghadam *et al.* [51] which enable the execution of queries over encrypted data (e.g., homomorphic encryption, property-preserving encryption) without decryption using analytics are both non-trivial and costly in terms of computing power and analytic processing difficulties.

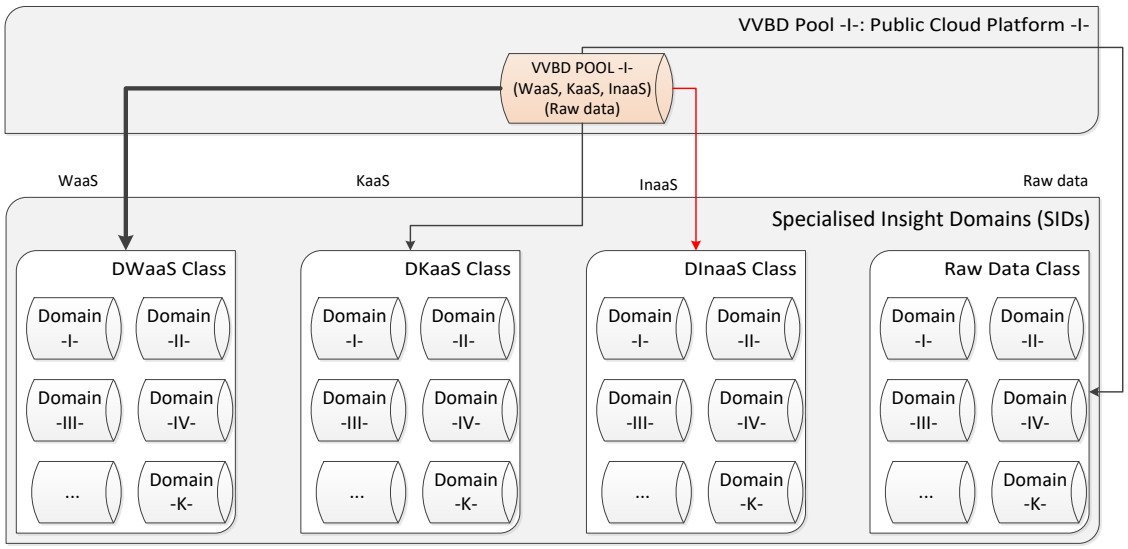

Figure 4: Organization of the insights and raw data in specialized domains in regard to the VVBD Pool.

An effective approach in which insights and sanitized, cleaned and filtered data are integrated, trustworthy, and efficiently accessible is presented in Section 4 by which cyberattacks can be mitigated and bigger data can be shared in public domains to help bridge the information gaps. In this framework proposal, the way of processing BD by service providers in an automated manner before placing in public domains is illustrated in Fig. 3. In this model, first, cloud service providers in collaboration with the owners categorize the data into three groups as 1) highly sensitive, 2) sensitive, and 3) non-sensitive. These three datasets are compiled as follows: 1) highly sensitive data usually encrypted is not sanitized and not placed in the VVBD pool in any case. Rather, it is analyzed within the cloud platform by BD analytics to extract insights such as WaaS, KaaS, and InaaS. These gained insights are placed in the VVBD pool, 2) sensitive data usually involving personal information is sanitized, filtered and analyzed by analytics to extract insights. Then, the extracted insights along with the anonymous filtered raw data are placed in the VVBD pool, and 3) non-sensitive data is placed in the VVBD pool after cleaned and filtered without any sanitization process and the insights gained from this data are also placed in the VVBD pool. The process of sanitization is mainly carried out for privacy protection where there is a trade-off between privacy and utility, which was discussed in [72] and [5]. Sanitization process may result in successful outcomes where effective tools developed for specific domains are utilized, an example of which is analyzed with medical domains by the study [73]. In addition to sanitization process, erroneous and noisy data is cleaned, and unnecessary, unhelpful data is filtered out to reduce the amount of BD substantially in highly summarized form in such a way that useful raw data that may result in insights is not discarded and lots of memory space and processing time is saved. The harvested insights can be used on a pay-as-you-go basis and more importantly can be combined with other harvested data or raw data in order to gain more insights as explored in Section 4 in detail.

The VVBD pool is maintained separately from the cloud's operational locations to make its use easier regarding the specific rules of authentication and automatization, and consequently to increase its availability with high performance focusing on prompt decision-making modeling. Sanitization of the data categorized as sensitive and the placement of raw data along with the sanitized data in the VVBD pool should be performed during real-time data streaming autonomously. Prioritization of the processing of insight extraction should be regulated according to the latency requirements regarding pay-as-you-go demands based on the data transfer and update cycles. Placement of the raw data and the extracted insights should be carried out based on the insight categorization and SIDs as explained in Section 3.2 in detail.

3.2. Organization of insights and raw data within specialized Insight Domains (SIDs) in VVBD Pools

Insights and filtered, sanitized and cleaned raw data or pure raw data obtained as illustrated in Fig. 3 are ag-

gregated into four classes in VVBD pools, namely Deep WaaS (DWaaS), Deep KaaS (DKaaS), Deep InaaS (DInaaS) and Raw Data as illustrated in Fig. 4. More explicitly, the placement is carried out in a consolidated way as all WaaSs in DWaaS class, all KaaSs in DKaaS class, all InaaSs in DInaaS class and all the remaining filtered, sanitized and cleaned raw data in Raw Data class in their SIDs. In other words, the pieces of the processed BD and VBD as explained in Section 3.1 are labeled and placed in dedicated labeled domains and sub-domains with sufficiently fine granularity. A domain corresponds to a specific field such as medicine and it is established in all the aforementioned four classes to enclose WaaSs, KaaSs, InaaSs, and Raw Data labeled with its name such as the medical domain. A domain may have various sub-domains to enclose the insights related to sub disciplines as illustrated in Fig. 5 for the medical domain. Domains and sup-domains may also be created on a subject-oriented and time-oriented basis. For instance, InaaS, KaaS, WaaS and raw data related to the medicine should be in the medical domain and the medical domain should consist of sub domains such as cancer, psychology, anatomy, genetics, etc. Furthermore, specialized sub-sub-domains within sub-domains increase the efficient and effective use of FCP. Most of the disciplines have similar agreed upon granularity standards, just changing slightly. The AIA is analyzed in Section 5 in the context of tackling privacy and security concerns while analyzing data to mine insights. New insights can be created using the current insights and raw data in these domains by the service provider autonomously on a pay-as-you-go manner. Furthermore, new insights can be generated by anybody, institutions, organizations, government and companies using AIA and the infrastructure established within FCP as explained in Sections 4 and 5.

4. Forged Cloud Platforms (FCP)

Data fusion is a technique to make an overall sense of data from different sources that commonly have different data structures [46]. Due to heavy network traffic with respect to data migrations across different datacenters, the underlying network infrastructure may not be able to transfer data packets from source to destination, resulting in performance degradation [8]. It gets worse when huge BD transformation is required across multiple cloud platforms based on bandwidth, storage, and computation limitations. The need for inter-datacentre migrations to handle BD processing increases the load on network infrastructure in multiple cloud platforms [8]. The current approaches are not

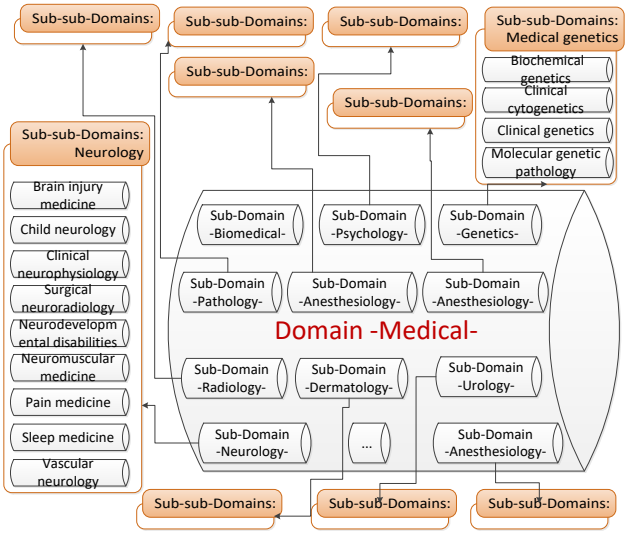

Figure 5: Structure of sub-domains within domains.

viable for managing VVVBD traffic in multiple cloud platforms and processing this gigantic BD to gain insights is not an easy process and feasible, mainly for low-latency applications as mentioned earlier. In this regard, in order to alleviate the shortcomings of the current architecture and concepts, FCP with an architecture of leveraging deep-learning from the insights created within multiple cloud platforms is designed to manage distributed insights effectively and efficiently within a cloud platform and across various cloud platforms in a virtual way of fusion of insights by 1) globally optimized resource allocation and orchestration, scheduling and insight distribution on a proactive user-centric paradigm and pay-as-you-go basis within reduced operation complexity, 2) minimizing data traffic substantially, 3) increasing response time to meet the requirements of low-latency applications, and 4) mitigating the privacy and security concerns. More explicitly, a conceptual VVVBD pool on FCP is built to enable autonomous enhanced insight generation and decision-making by coordinating, combining, integrating and distributing deep wisdom, knowledge, information and processed raw data in a concise view ensuring consistency. This framework is designed to provide uninterrupted and quick access to pre-created accumulated insights and create new insights out of the existing pre-created and aggregated insights from multiple virtual sources in an organized way with several levels of granularity on a subject-oriented and time-oriented basis.

In the framework of FCP illustrated in Fig. 6, VVBD Pools are physically separated from the main cloud platforms and updated continuously and autonomously by

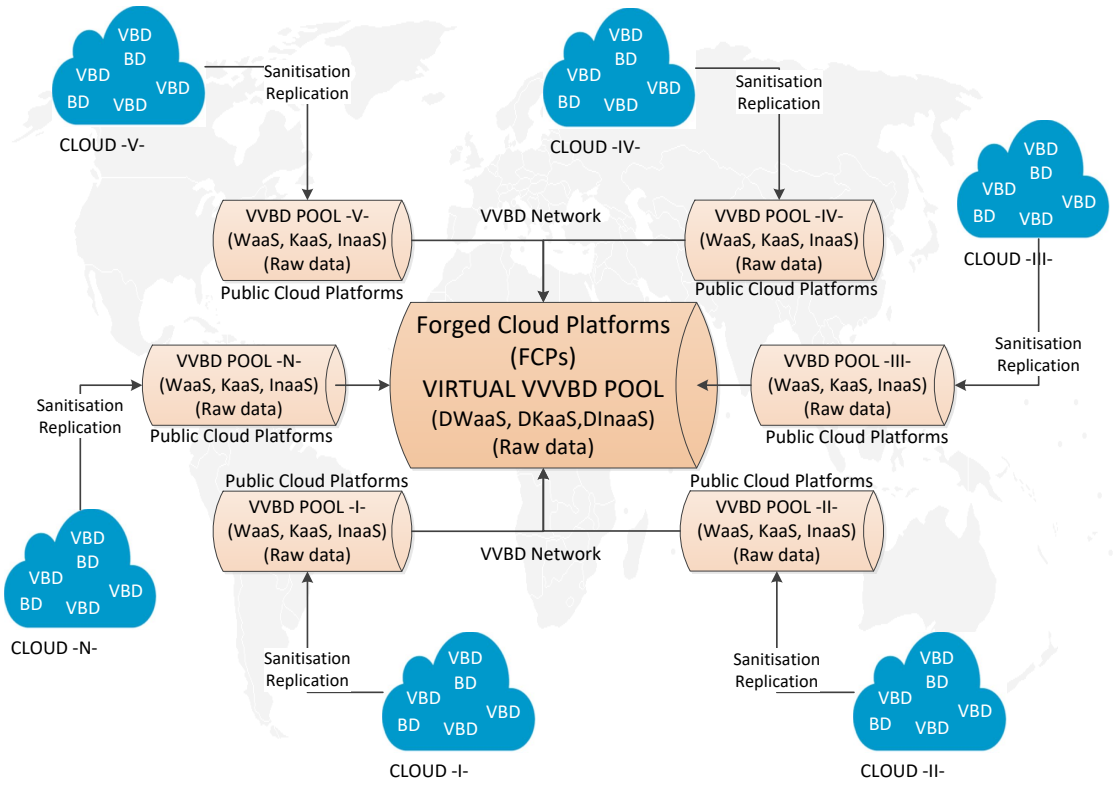

Figure 6: Framework of Forged Cloud Platforms (FCP): Very Very Very BD (VVVBD) pool.

the service providers to tackle the outdated insights and data on a near-real-time basis. FCP, in other words, virtual VVVBD pool in which there is no highly sensitive information is formed from these VVBD pools and open to everybody with relaxed authorization and authentication rules and regulations for extracting business insights with instrumental abilities. The main components of FCP framework are presented in Fig. 7. The FCP architecture incorporates the ability into a cloud environment to share the globally created data and insights among multiple cooperating users by mitigating the privacy and security concerns. Sufficiently and properly granulated organized and structured deep insight domains and related sub-domains should be established within the VVVBD pool in a more consolidated way to integrate and aggregate the 4 classes — DWaaS, DKaaS, DInaaS, and Raw Data mentioned in Section 3.2 within effective addressable virtual mechanisms. The component of DWaaS classes contains the addresses of DWaaSs in specific domains established in the distributed cloud platforms as explained in Section 3.2. Similarly, the other components, namely, DKaaS, DInaaS, and Raw Data maintain the related addresses as well. Harnessing the ready-to-use insights will help reduce the data traffic on the network backbone significantly. These deep insights can be on a pay-as-you-go basis as an incentive to inspire the owners of BD and cloud companies to take part in this framework voluntarily by which a tremendous wealth will be created in this structure benefiting all the stakeholders. The insights acquired or created in FCP can be an input to IoT and/or AMSs, or they can be merged with other insights created locally for better decision-making and actuation as illustrated in Fig. 8. With FCP, insights are not only gained and organized in dedicated domains for current users' needs, but also, likely users' future service requests are predicted proactively to direct the enterprises to follow an intelligent way of doing their businesses.

Sending the entire datasets across the extreme ends in the infrastructure is unrealistic and the approaches that collect data and perform computational processing near the data source itself present a more practical and realistic alternative [74] due to large volumes of data produced every minute and bandwidth limitations. Analyzing data at the early stages of infrastructure pipeline presents additional benefits of data and communication security in the overall system, owing to the fact that raw data is now processed closer to the data source, and only

processed data is sent further [74]. In this sense, in order to reduce the workload burden on the cloud and network, processing of BD generated at the edge platforms locally with respect to the domain and insight structure mention in Section 3.2 using effective AIA tools (e.g., advanced hybrid edge-cloud analytics) is crucially important in order to produce insights at the edges to be submitted to the cloud platform along with sanitized, cleaned and filtered data rather than unprocessed raw data as shown in Fig. 8. In this way, processing of BD starts from its source, which is immensely valuable for the cloud and will ease the later required data processing phases mentioned throughout this paper in regard to FCP by reducing both resource requirements on the related cloud platforms and on FCP, and energy consumption at the network usage.

FCP coupled with powerful edges enables a simple virtual layer design and effectively bridges the latency gap between BD cloud computing and real-time network optimization. With FCP, energy efficiency is maximized within an end-to-end (E2E) communication, rather, the processing burden is mounted on the main cloud datacenters and edge platforms by which low latency requirements and prompt decision-making abilities are provided for low-latency applications with ready-to-use abilities, mainly for subscribes on a pay-as-you-go basis. Ultra-low-latency is one of the major requirements of 5G RANs [75]. There are applications requiring ultra-low-latency in real-time such as intelligent transportation systems (ITSs). Such data analytics capabilities cannot be provided by conventional cloud centric data processing techniques whose communication and computing latency can be high [76]. The proposed FCP in this study with ready-to-use insights along with cloud platforms with cloudlets. Experiments show that the use of cloudlets decreases response time by 51% and reduces energy consumption by up to 42% in a mobile device compared to cloud offload [16]. approaching to the users and cutting edge communication technologies (e.g., 5G and beyond) propose an effective design architecture within a novel concept to support such systems in real-time as illustrated in Fig. 8, supported by local sensors at the edge/fog platforms as in ITSs edge platform where every vehicle can be equipped with sensors and capable processing devices and smart analytics [76].

With FCP using AIA, insights are not only gained and organized in dedicated domains for current users' needs, but also, likely users' future service requests are predicted proactively to direct the enterprises to follow an intelligent way of doing their businesses.

5. Advanced Insight Analytics (AIA) and Deep Insight as a Service (DINSaaS)

There is an exponentially growing gap between the generation of data and the benefit we get from it [7] as data piles up exponentially. A recent research by IBM shows that 1% of data collected by organizations is used for analysis [77]. Limits to what can be learned from BD often boil down to how much data can be feasibly and economically processed [9]. The approach proposed in this paper is designed to tackle this issue effectively and efficiently by gaining insights from specific public domains as explained in Section 3.2 and illustrated in Figs. 4 and 7. One of the main features of the BD analytics, particularly DM tools, should be taking care of privacy and security while driving insights from BD by engaging with cybersecurity abilities. In the approach, the problem space is decreased significantly with pre-created insights and sanitized and filtered raw data. In this decreased complexity, the objective of competent and scalable AIA is to extract new advanced insights, in other words, DINSaaS, from pre-created insights along with filtered, sanitized and cleaned raw data within a specific domain/domains and subdomains through well-directed predictions using distributed and parallel scalable algorithms to make overall decision-making process easier and to improve the quality of life. Several ensemble ML models (e.g., boosting, bagging and Stacking) can combine the results of several techniques and make a final better decisions [7]: Boosting (e.g., AdaBoost) uses voting (for classification) or averaging (for numeric prediction) to combine the output of individual models of the same type (e.g., only decision trees), weights a model's contribution by its performance rather than giving equal weight to all models and weighting is used to give more influence to the more successful ones; Bagging combines models of the same type (e.g., only decision trees) and the models receive equal weight; Stacking is applied to models built by different learning algorithms such as decision tree inducer, a Naive Bayes learner, and an instance-based learning method and all three are used for prediction and combine the outputs together by voting; in this way, Stacking tries to learn which classifiers are the reliable ones, using another learning algorithm-the metalearner-to discover how best to combine the output of the base learners. Besides these similar approaches, new AIA techniques and tools are required to be developed specific to the characteristics of the domains partitioned in FCP. More explicitly, new AIA techniques are needed to analyze the insights placed in four classes and various dedicated domains as explained in Section 3.2 ef-

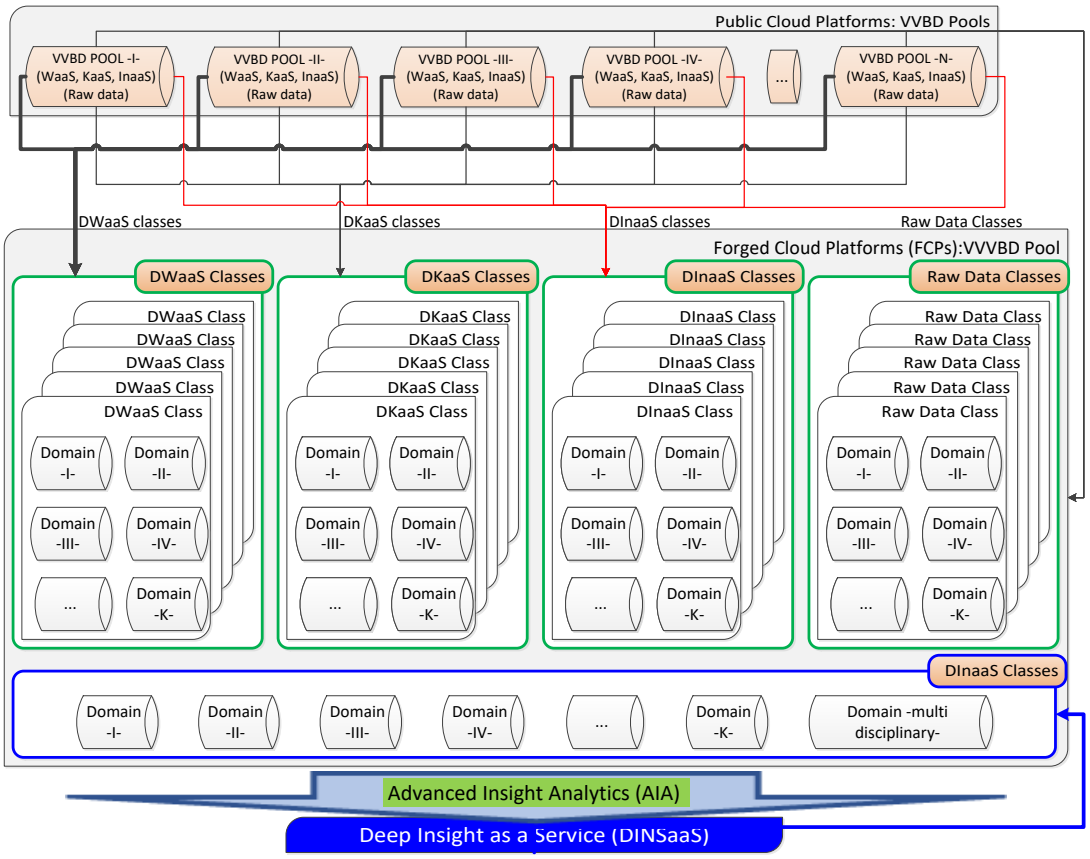

Figure 7: Framework of the Deep Insight as a Service (DINSaaS) on Forged Cloud Platforms (FCP).

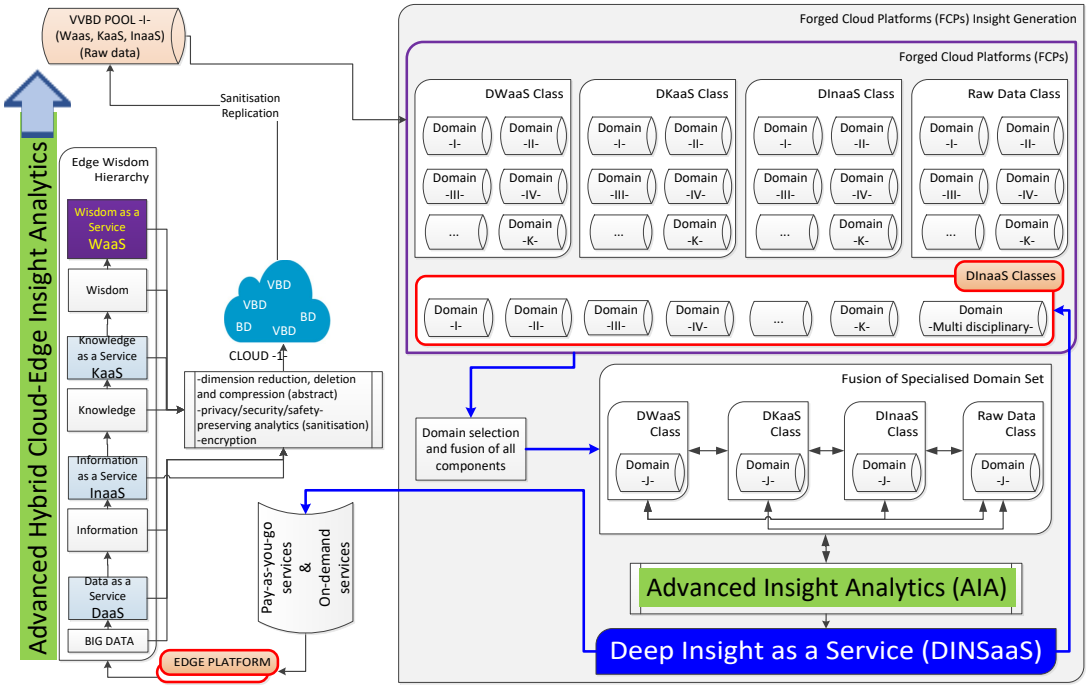

Figure 8: Framework of effective processing of AIA and interaction of the edge/fog and FCP.

fectively using the integration, analysis, and interpretation in order to make sense out of the overwhelming amount of data by exploiting DWaaS, DKaaS and DInaaS along with the sanitized and filtered raw data with reduced latency requirements. The characteristics and entities of WaaS, KaaS, and InaaS within the classes in domains (Fig. 7) should be specified based on the data features of 5 V's and their data signature — veracity, volume, value, variety and velocity (e.g., occurrence time, update frequency) in addition to duration validity, geographical occurrence, resources, etc in order to enable better decision making. We will be analyzing and presenting similar AIA techniques in the following study due to the page limits here. In that study, we will be focusing the weights of the pre-created insights, in other words, the proportion of wisdom contribution in decision-making process based on the characteristics of the insights and BD from which they are derived.

Particular features of the AIA can be summarized as follows:

1) AIA tools should be supported with powerful ML, Deep Learning (DL), Reinforcement Learning (RL), DM techniques.

2) AIA tools should have effective fusing abilities and be able to incorporate the data signatures of pre-created insights into decision-making process in order to weight the outcomes of the insights appropriately with respect to the characteristics of sources from which they are generated.

3) The privacy and security issues should be tackled by AIA very well and they can be employed as a gateway to the sanitized and filtered sensitive data within specific domains. In this way, direct access to sensitive raw data may be restricted and AIA can be permitted to run by mitigating privacy and security concerns while extracting insights.

4) Synergistic integration of various domains at a time may be required to be processed by AIA in order to extract much better insights. Therefore, AIA should be able to perform on multiple domains in FCP at a time. For instance, load forecasting can be managed effectively in the electricity supply industry within the smart grid where it is crucial to determine future demand for power as far in advance as possible. If accurate estimates can be made for the maximum and minimum load for each hour, day, month, season, and year, utility companies can make significant economies in areas such as setting the operating reserve, maintenance scheduling, and fuel inventory management. These similar decision-making abilities require the analysis of multiple domains in parallel processing using multiple threads.

6. Challenges

The emerging BD pose new challenges, some of which cannot be adequately addressed by existing infrastructure, data analytics, state-of-the-art cybersecurity solutions, cloud and host computing models alone, which exacerbates the current situation with respect to exponentially increasing voluminous BD. To unleash DInaaS potential, these challenges raising concerns must be addressed well in terms of both technical infrastructure and the management of data, and human factors, such as privacy and security. The main challenges are presented as follows from the most crucial ones:

1) Concrete agreement difficulties among cloud service providers: Agreement between parties on many issues is not easy and this process will effect how regulation and legislation will evolve.

2) Integration of smart domains: Main actors as decision-makers in smart domains avoid sharing their data due to serious security and privacy concerns, which reduces the further integration of the smart domains even though the number of successful examples and attempts of combining these domains is increasing.

3) Reluctance in data sharing: Data sharing between prominent leading companies that generate BD is highly limited regarding privacy and security concerns, insufficient rules and regulations, and competitive and commercial risks. Additionally, data owners do not volunteer to share their data particularly due to the breaches of their privacy and security. Effective incentive mechanisms along with effective privacy and security preserving tools should be developed and employed to encourage data sharing.

4) National/international rules and regulations: The legal challenges, in particular, about the privacy of people using cyber-physical domains have yet to be solved [78] [3]. The responsibilities of cloud service providers against the customers are not well defined within the national and international laws of electronic commerce.

5)Management of communication with respect to emerging future networks: The studies on the management of large volumes of BD traffic regarding emerging communication technologies (e.g., 5G and beyond) are not adequate to deploy these technologies effectively. A recent paper [79] discusses how to manage the E2E traffic with recently emerging communication technologies

[3]The motivated readers are referred to [78] for the analysis of privacy and security legislation regarding IoT and cloud use.

effectively. More experimental studies and more real-world examples are needed in this field in order to be able to deploy these communication technologies in an efficient and effective manner regarding very BD.

6) Energy consumption and optimization concerns: Datacenters cost hundreds of millions per year and consume more than 61 billion kilowatt-hours, one-fifth percent of the country's entire energy consumption [34]. Our approach loads an extra workload and burden on cloud platforms regarding more energy consumption and more storage units even though it is reduced on the network with much less traffic and at the edges with much less processing requirements. Advanced optimization and advanced data analytics tools are needed to be built to increase the efficacy of these platforms for environmental benefits and cost reduction; an attempt for which iSpot is developed to be able to guarantee the performance of BD analytics running on cloud transient servers while reducing the job budget by up to 83.8% in comparison to the state-of-the-art server provisioning strategies, yet with acceptable runtime overhead [44]. Another attempt to run the BD analytics in geo-distributed datacenters effectively can be found in the study [48]. Furthermore, software-defined-network-based optimization of BD management across multi-cloud datacenters was analyzed in [8].

7) Lack of standards: Agreed upon standards and protocols for effective policies and mechanisms are necessary. The General Data Protection Regulation (GDPR) and the ongoing e-privacy regulation effort are significant steps in regulating the protection of sensitive information by placing obligations on data controllers and data processors, as well as specifying user's rights. However, no specific algorithms are mentioned, and hence we are far from effective standardization guidelines [56]. The complexity of secure data analysis will require several kinds of standards, related not only to the different aspects of privacy-preserving analytics, but also related issues like personal data management and consent [56].

8) Insufficient tools for harvesting insights: AIA tools that can acquire insights within the structure of FCP will be needed along with efficient resource orchestration, insight and content distribution enabling audits to ensure compliance. These tools should embrace the text mining, web mining, speech mining, image/video mining abilities and should be able to tackle the unstructured BD very well since most of the BD is in unstructured format, enabling discovering intrinsic relationships, text clustering, text categorization, concept/entity extraction, production of granular taxonomies, document summarization, sentiment analysis, entity relation modeling.

9) Sanitization difficulties: The need for robust privacy-preserving data analysis technologies has been recognized by both regulators and industry [56]. New effective privacy-protective techniques and mechanisms are needed to retain privacy when analyzing users' data and these mechanisms should be active in AIA. The moral is that if you really do remove all possible identification information from a database, you will probably be left with nothing useful [7]. It takes a long time to sanitize BD and BD may lose its meaning with the side effects of the sanitization process. A fully-fledged approach to privacy-preserving data analysis would still require significant interdisciplinary effort, some of which have to do with issues such as effective personal data management and consent [56].

10) Lack of hybrid cloud-edge analytics at the edge/fog platform: Data is subject to attacks and security breaches when stored in the cloud platform. Effective AIA at the edges producing insights to be submitted to the cloud platform along with sanitized and filtered data rather than unprocessed raw data is extremely crucial in our design to reduce the workload burden on the cloud and network and to improve the overall efficacy of the architecture.

7. Discussion

Cheap ubiquitous computing enables the collection of massive amounts of personal data in a wide variety of domains [80]. In cloud platforms, most of the time, BD can not be reached because of the privacy and security concerns and effective tools are being deployed to detect and respond faster to cyberthreats, attacks, breaches of data. One of the recent popular trends is to integrate all smart domains in a combined architecture of the cloud platform [81] to create bigger synergies even though it involves many challenges.

In future mobile networks, such as 5G - e.g., the Radio-Access Network (RAN), emerging smart services are expected to support billions of smart devices with unique characteristics and traffic patterns [82] to facilitate the application of wireless BD and to achieve a flexible and efficient communication, consequently an excellent synergy using the smart platforms and wiser domains. High-level topics concerning today's production of goods and services include sustainability, flexibility, efficiency, and competitiveness [83]. In this manner, today's rapid changing technological and business environment urges the companies to be agile in order to adapt to upheaval market fluctuations, cope with unprecedented threats and most importantly thrive in a

competitive business environment, even during recessions by foreseeing and exploiting the emerging business opportunities. Following the frameworks proposed in this study, the exponentially increasing volumes of data will turn into insights using advanced data analytics, which, in turn, will lead to an aggregation of wisdom to optimize all processes to ensure higher quality services and goods are manufactured at a lower cost. More explicitly, following this framework, transforming the business and products into more intelligent, more autonomous services and products with increased customizable functionalities will be possible to maintain a competitive edge by meeting the market dynamics, in particular, changing consumer habituation and demands. Moreover, DINSaaS will bring huge economic value for the wiser IT industry based on the instant pay-as-you-go services. The proposed framework, FCP, in which too many smart ideas to be explored and exploited will provide a wiser environment and particularly, DInaaS will be the core architecture of BI in the coming age of the cyber-physical world.

To prevent chaos in the hyper-connected world, businesses need to make every effort to reduce the complexity of connected systems, enhance the security and standardization of applications, and guarantee the safety and privacy of users anytime, anywhere, on any device [84]. By focusing on user interaction and configurability, lifetime optimization, intelligent analysis of BD, location independent monitoring and control, data security and reduced system complexity with FCP using effective management of gained insights, the way of doing business more intelligently will be realized within AoE.

The framework proposed here requires the rules and regulations to be amended by the national/international leading organizations and governments with the help of the leading companies, particularly, IT-based companies. Establishment of effective sanitization techniques, tools and approaches would enable opportunities of extraction insights from highly sensitive data based on effective regulations and standardization, and additionally, publishing insights rather than raw data would immensely serve sanitization purposes further, which would, in turn, accelerate data sharing and trigger better and more insight generation in public domains.

The management of cloud datacenters scattered around the globe supported by cloudlets and large network webs is extremely expensive. The main cloud service providers are expected to invest an amount of $383 billion into their cloud infrastructures in 2020 [85]. In this regard, leading competitive cloud service providers such as Google, Amazon, Microsoft, IBM will merge their powers and marry under bigger joint ventures not only to reduce the current immense management costs and future investment costs substantially, but also to increase the efficacy of their services in a fruitful and prosperous way of exploiting many more datacenters and much larger global networks, particularly, by maximizing the resource utilization while mitigating the performance degradation. The aforementioned joint ventures will help realize the FCP architecture proposed in this study faster, with increasing data sharing abilities resulting in both increased productivity through the booms of insights gained from gigantic data, and effective way of doing business by intertwining smart platforms and businesses, particularly conducted by leading companies, which will definitely make our life cheaper, easier, more intelligent.

8. Conclusions and Future Research Ideas

To become and remain competitive, enterprises must seek to adopt advanced analytics, and adapt their business models, establish specialist data science teams and rethink their overall strategies to keep pace with the competition [43]. The transformation of smart domains and platforms into smarter domains with FCP presented in this paper will foster the development of industry to make our life better and simpler. More explicitly, by adopting the approaches proposed in this study, 1) bigger data traffic will be managed effectively and efficiently with less overload on the network backbone by trafficking insights rather than very big raw data, and bandwidth constraints will be mitigated, 2) reduced latency within reasonable responses will be enabled with near real-time processing of geo-distributed BD, 3) the chaos in the hyper-connected world will be mitigated with a systematic solution, 4) further insights will be generated within organized domains, 5) the future will be foreseen better and easier with effective decision-making decisions, and new invaluable opportunities will be revealed to be explored and exploited, 6) exponential increase of cyberattacks to highly sensitive information will be mitigated, 7) the structure of the services and advanced products will be less complex, consequently less error-prone with the use of the wiser inputs and consequently wiser systems, 8) the efficacy of smart products will increase through effective product life-cycle management within AoE, 9) the rapid product and service customization will be possible, 10) cost will decrease substantially with less number of sensors and robust insight inputs within less complex structures, 11) customer satisfaction will increase within smoothly working environment with 24/7 seamlessly working products, 12) new business models, improving efficiency and

increasing employee and customer engagement within AoE will be developed, 13) the invention of efficient and robust business models with FCP will further advance the future research and innovative products in a perpetual revolution of the industry.

The development of new AIA approaches mentioned in Section 5 will be carried out in a future study to be able to manage the future cloud insight architectures such as FCP proposed in this paper. The characteristics and entities of insights will be analyzed in the future study. Cryptographic schemes and practical systems which enable the execution of queries over encrypted data without decryption using analytics will be immensely focused both in order to mitigate the security, privacy and cybersecurity concerns, and in order to reduce the computation overhead caused by the encryption in the following years.

References

[1] J. Chen, J. Ma, N. Zhong, Y. Yao, J. Liu, R. Huang, W. Li, Z. Huang, Y. Gao, J. Cao, Waas: Wisdom as a service, IEEE Intelligent Systems 29 (6) (2014) 40–47. doi:10.1109/MIS.2014.19.

[2] A. C. Victor, S. Rao, Analytics on the cloud, IEEE Potentials 37 (4) (2018) 24–27. doi:10.1109/MPOT.2018.2824358.

[3] G. Shmueli, P. C. Bruce, I. Yahav, N. R. Patel, J. Kenneth C. Lichtendahl, Data Mining for Business Analytics: Concepts, Techniques, and Applications in R, 2nd Edition, Wiley, NJ, USA, 2018.

[4] K. Kuru, H. Yetgin, Transformation to advanced mechatronics systems within new industrial revolution: A novel framework in automation of everything (aoe), IEEE Access 7 (2019) 41395–41415. doi:10.1109/ACCESS.2019.2907809.

[5] Z. He, Z. Cai, J. Yu, Latent-data privacy preserving with customized data utility for social network data, IEEE Transactions on Vehicular Technology 67 (1) (2018) 665–673. doi:10.1109/TVT.2017.2738018.

[6] S.-S. Yau, H. An, A. Buduru, An approach to data confidentiality protection in cloud environments, International Journal of Web Services Research 9 (3) (2012) 67–83. doi:10.4018/jwsr.2012070104.

[7] I. H. Witten, E. Frank, M. A. Hall, C. J. Pal, Data Mining: Practical Machine Learning Tools and Techniques, 4th Edition, Morgan Kaufmann, Illinois, USA, 2016.

[8] R. Chaudhary, G. S. Aujla, N. Kumar, J. J. P. C. Rodrigues, Optimized big data management across multi-cloud data centers: Software-defined-network-based analysis, IEEE Communications Magazine 56 (2) (2018) 118–126. doi:10.1109/MCOM.2018.1700211.

[9] P. Eugster, C. Jayalath, K. Kogan, J. Stephen, Big data analytics beyond the single datacenter, Computer 50 (6) (2017) 60–68. doi:10.1109/MC.2017.163.

[10] S. Ji, B. Li, Wide area analytics for geographically distributed datacenters, Tsinghua Science and Technology 21 (2) (2016) 125–135. doi:10.1109/TST.2016.7442496.

[11] J. Chen, J. Han, Y. Deng, H. Zhong, N. Wang, Y. Li, Z. Wan, T. Kotake, D. Wang, N. Zhong, Wisdom as a service for mental health care, IEEE Transactions on Cloud Computing (2018) 1–1doi:10.1109/TCC.2015.2464820.

[12] X. Chen, L. Jiao, W. Li, X. Fu, Efficient multi-user computation offloading for mobile-edge cloud computing, IEEE/ACM Transactions on Networking 24 (5) (2016) 2795–2808. doi:10.1109/TNET.2015.2487344.

[13] M. Chiang, T. Zhang, Fog and iot: An overview of research opportunities, IEEE Internet of Things Journal 3 (6) (2016) 854–864. doi:10.1109/JIOT.2016.2584538.

[14] R. S. Montero, E. Rojas, A. A. Carrillo, I. M. Llorente, Extending the cloud to the network edge, Computer 50 (4) (2017) 91–95. doi:10.1109/MC.2017.118.

[15] M. Satyanarayanan, R. Schuster, M. Ebling, G. Fettweis, H. Flinck, K. Joshi, K. Sabnani, An open ecosystem for mobile-cloud convergence, IEEE Communications Magazine 53 (3) (2015) 63–70. doi:10.1109/MCOM.2015.7060484.

[16] Y. Ai, M. Peng, K. Zhang, Edge computing technologies for internet of things: a primer, Digital Communications and Networks 4 (2) (2018) 77 – 86. doi:https://doi.org/10.1016/j.dcan.2017.07.001.
URL http://www.sciencedirect.com/science/article/pii/S2352864817301335

[17] M. Gharbaoui, B. Martini, D. Adami, S. Giordano, P. Castoldi, Cloud and network orchestration in sdn data centers: Design principles and performance evaluation, Computer Networks 108 (2016) 279 – 295. doi:https://doi.org/10.1016/j.comnet.2016.08.029.
URL http://www.sciencedirect.com/science/article/pii/S1389128616302821

[18] T. Chen, X. Gao, G. Chen, The features, hardware, and architectures of data center networks: A survey, Journal of Parallel and Distributed Computing 96 (2016) 45 – 74. doi:https://doi.org/10.1016/j.jpdc.2016.05.009.
URL http://www.sciencedirect.com/science/article/pii/S0743731516300399

[19] B. Wang, Z. Qi, R. Ma, H. Guan, A. V. Vasilakos, A survey on data center networking for cloud computing, Computer Networks 91 (2015) 528 – 547. doi:https://doi.org/10.1016/j.comnet.2015.08.040.
URL http://www.sciencedirect.com/science/article/pii/S138912861500300X

[20] O. Consortium, Openfog architecture overview (2016) [cited 25.12.2018].
URL https://www.alibabacloud.com/blog/bringing-iot-to-the-cloud-fog-computing-and-cloudlets_593824

[21] Z. Sheng, S. Yang, Y. Yu, A. V. Vasilakos, J. A. Mccann, K. K. Leung, A survey on the ietf protocol suite for the internet of things: standards, challenges, and opportunities, IEEE Wireless Communications 20 (6) (2013) 91–98. doi:10.1109/MWC.2013.6704479.

[22] D. Miorandi, S. Sicari, F. D. Pellegrini, I. Chlamtac, Internet of things: Vision, applications and research challenges, Ad Hoc Networks 10 (7) (2012) 1497 – 1516. doi:https://doi.org/10.1016/j.adhoc.2012.02.016.
URL http://www.sciencedirect.com/science/article/pii/S1570870512000674

[23] K. Rose, S. D. Eldridge, L. Chapin, The internet of things : An overview understanding the issues and challenges of a more connected world, 2015.

[24] E. Borgia, The internet of things vision: Key features, applications and open issues, Computer Communications 54 (2014) 1 – 31. doi:https://doi.org/10.1016/j.comcom.2014.09.008.
URL http://www.sciencedirect.com/science/article/pii/S0140366414003168

[25] Q. Zhang, Z. Yu, W. Shi, H. Zhong, Demo abstract: Evaps:

Edge video analysis for public safety, in: 2016 IEEE/ACM Symposium on Edge Computing (SEC), 2016, pp. 121–122. doi:10.1109/SEC.2016.30.

[26] Y. Lee, W. Hsiao, Y. Lin, S. T. Chou, Privacy-preserving data analytics in cloud-based smart home with community hierarchy, IEEE Transactions on Consumer Electronics 63 (2) (2017) 200–207. doi:10.1109/TCE.2017.014777.

[27] IDC, The internet of things (2014) [cited 25.03.2019]. URL https://www.emc.com/leadership/digital-universe/2014iview/internet-of-things.htm

[28] K. A. Beaty, J. M. Chow, R. L. F. Cunha, K. K. Das, M. F. Hulber, A. Kundu, V. Michelini, E. R. Palmer, Managing sensitive applications in the public cloud, IBM Journal of Research and Development 60 (2-3) (2016) 4:1–4:13. doi:10.1147/JRD.2015.2513720.

[29] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, M. Ayyash, Internet of things: A survey on enabling technologies, protocols, and applications, IEEE Communications Surveys Tutorials 17 (4) (2015) 2347–2376. doi:10.1109/COMST.2015.2444095.

[30] DOMO, Data never sleeps 7.0 (2019) [cited 25.08.2019]. URL https://www.domo.com/learn/data-never-sleeps-7

[31] DOMO, Data never sleeps 6.0 (2018) [cited 25.01.2019]. URL https://www.domo.com/learn/data-never-sleeps-6

[32] K. Abouelmehdi, A. Beni-Hessane, H. Khaloufi, Big healthcare data: preserving security and privacy, Journal of Big Data 5 (1) (2018) 1. doi:10.1186/s40537-017-0110-7. URL https://doi.org/10.1186/s40537-017-0110-7

[33] M. J. Kaur, P. Maheshwari, Building smart cities applications using iot and cloud-based architectures, in: 2016 International Conference on Industrial Informatics and Computer Systems (CIICS), 2016, pp. 1–5. doi:10.1109/ICCSII.2016.7462433.

[34] Y. Sun, H. Song, A. J. Jara, R. Bie, Internet of things and big data analytics for smart and connected communities, IEEE Access 4 (2016) 766–773. doi:10.1109/ACCESS.2016.2529723.

[35] J. Gubbi, R. Buyya, S. Marusic, M. Palaniswami, Internet of things (iot): A vision, architectural elements, and future directions, Future Generation Computer Systems 29 (7) (2013) 1645 – 1660, including Special sections: Cyber-enabled Distributed Computing for Ubiquitous Cloud and Network Services and Cloud Computing and Scientific Applications — Big Data, Scalable Analytics, and Beyond. doi:https://doi.org/10.1016/j.future.2013.01.010. URL http://www.sciencedirect.com/science/article/pii/S0167739X13000241

[36] S. Tayeb, S. Latifi, Y. Kim, A survey on iot communication and computation frameworks: An industrial perspective, in: 2017 IEEE 7th Annual Computing and Communication Workshop and Conference (CCWC), 2017, pp. 1–6. doi:10.1109/CCWC.2017.7868354.

[37] M. Amadeo, C. Campolo, A. Iera, A. Molinaro, Information centric networking in iot scenarios: The case of a smart home, in: 2015 IEEE International Conference on Communications (ICC), 2015, pp. 648–653. doi:10.1109/ICC.2015.7248395.

[38] J. Han, M. Kam, J. Pei, Data Mining Concepts and Techniques,, 3rd Edition, Elsevier, San Francisco, USA, 2012.

[39] W. Yu, F. Liang, X. He, W. G. Hatcher, C. Lu, J. Lin, X. Yang, A survey on the edge computing for the internet of things, IEEE Access 6 (2018) 6900–6919. doi:10.1109/ACCESS.2017.2778504.

[40] J. K. Winn, B. Wright, The Law of Electronic Commerce, 4th Edition, Wolters Kluwer, NY, USA, 2019.

[41] J. Chen, J. Ma, N. Zhong, Y. Yao, J. Liu, R. Huang, W. Li, Z. Huang, Y. Gao, Waas—wisdom as a service, in: N. Zhong, J. Ma, J. Liu, R. Huang, X. Tao (Eds.), Wisdom Web of Things, Springer Nature, Springer International Publishing Switzerland, 2016, Ch. 2, pp. 27–46.

[42] Gartner, Gartner data shows 87 percent of organizations have low bi and analytics maturity (2019) [cited 12.06.2018]. URL https://www.gartner.com/en/newsroom/press-releases/2018-12-06-gartner-data-shows-87-percent-of-organizations-have-low-bi-and-analytics-maturity

[43] D. Laney, A. Jain, 100 data and analytics predictions through 2021 (2018) [cited 25.08.2019]. URL https://www.gartner.com/en/doc/3746424-100-data-and-analytics-predictions-through-2021

[44] F. Xu, H. Zheng, H. Jiang, W. Shao, H. Liu, Z. Zhou, Cost-effective cloud server provisioning for predictable performance of big data analytics, IEEE Transactions on Parallel and Distributed Systems 30 (5) (2019) 1036–1051. doi:10.1109/TPDS.2018.2873397.

[45] B. Marr, Big data: using SMART big data. analytics and metrics to make better decisions and improve performance, 5th Edition, Wiley, Chichester, UK, 2015.

[46] Y. He, F. R. Yu, N. Zhao, H. Yin, H. Yao, R. C. Qiu, Big data analytics in mobile cellular networks, IEEE Access 4 (2016) 1985–1996. doi:10.1109/ACCESS.2016.2540520.

[47] G. T. Lakshmanan, R. Khalaf, Leveraging process-mining techniques, IT Professional 15 (5) (2013) 22–30. doi:10.1109/MITP.2012.88.

[48] Q. Xia, W. Liang, Z. Xu, Data locality-aware big data query evaluation in distributed clouds, The Computer Journal 60 (6) (2017) 791–809. doi:10.1093/comjnl/bxw101.

[49] Y. Sun, H. Yan, J. Zhang, Y. Xia, S. Wang, R. Bie, Y. Tian, Organizing and querying the big sensing data with event-linked network in the internet of things, International Journal of Distributed Sensor Networks 10 (8) (2014) 218521. arXiv:https://doi.org/10.1155/2014/218521, doi:10.1155/2014/218521. URL https://doi.org/10.1155/2014/218521

[50] S. K. Sharma, X. Wang, Live data analytics with collaborative edge and cloud processing in wireless iot networks, IEEE Access 5 (2017) 4621–4635. doi:10.1109/ACCESS.2017.2682640.

[51] S. Sobati Moghadam, A. Fayoumi, Toward securing cloud-based data analytics: A discussion on current solutions and open issues, IEEE Access 7 (2019) 45632–45650. doi:10.1109/ACCESS.2019.2908761.

[52] GALLUP, Americans: Credit card information still getting hacked (2016) [cited 27.10.2016]. URL https://news.gallup.com/poll/196802/americans-credit-card-information-getting-hacked.aspx

[53] L. H. Newman, Hack brief: Hackers breach a billion yahoo accounts. a billion (2016) [cited 14.12.2016]. URL https://www.wired.com/2016/12/yahoo-hack-billion-users/

[54] ICloudPE, How cloud storage became a target for hackers – and what can be done about it (2016) [cited 20.10.2016]. URL https://icloud.pe/blog/how-cloud-storage-became-a-target-for-hackers-and-what-can-be-done-about-it/

[55] P. O. Obitade, Big data analytics: a link between knowledge management capabilities and superior cyber protection, Journal of Big Data 6 (1) (2019) 71. doi:10.1186/s40537-019-0229-9. URL https://doi.org/10.1186/s40537-019-0229-9

[56] J. Crowcroft, A. Gascón, Analytics without tears or is there a way for data to be anonymized and yet still useful?, IEEE Internet Computing 22 (3) (2018) 58–64. doi:10.1109/MIC.2018.032501518.

[57] L. Arbuckle, K. E. Emam, Anonymizing Health data, 1st Edition, O'Reilly Media, Inc., Surrey, UK, 2013.

[58] L. SWEENEY, k-anonymity: A model for protecting privacy, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 10 (05) (2002) 557–570. arXiv:https://doi.org/10.1142/S0218488502001648, doi:10.1142/S0218488502001648.
URL https://doi.org/10.1142/S0218488502001648

[59] C. Dwork, Differential privacy, in: M. Bugliesi, B. Preneel, V. Sassone, I. Wegener (Eds.), Automata, Languages and Programming, Springer Berlin Heidelberg, Berlin, Heidelberg, 2006, pp. 1–12.

[60] A. Machanavajjhala, J. Gehrke, D. Kifer, M. Venkitasubramaniam, L-diversity: privacy beyond k-anonymity, in: 22nd International Conference on Data Engineering (ICDE'06), 2006, pp. 24–24. doi:10.1109/ICDE.2006.1.

[61] C. Liu, P. Mittal, Linkmirage: Enabling privacy-preserving analytics on social relationships, in: NDSS, 2016.

[62] J. C. Lin, J. M. Wu, P. Fournier-Viger, Y. Djenouri, C. Chen, Y. Zhang, A sanitization approach to secure shared data in an iot environment, IEEE Access 7 (2019) 25359–25368. doi:10.1109/ACCESS.2019.2899831.

[63] Z. Cai, Z. He, X. Guan, Y. Li, Collective data-sanitization for preventing sensitive information inference attacks in social networks, IEEE Transactions on Dependable and Secure Computing 15 (4) (2018) 577–590. doi:10.1109/TDSC.2016.2613521.

[64] D. Sánchez, M. Batet, A. Viejo, Automatic general-purpose sanitization of textual documents, IEEE Transactions on Information Forensics and Security 8 (6) (2013) 853–862. doi:10.1109/TIFS.2013.2239641.

[65] Y. Hong, J. Vaidya, H. Lu, P. Karras, S. Goel, Collaborative search log sanitization: Toward differential privacy and boosted utility, IEEE Transactions on Dependable and Secure Computing 12 (5) (2015) 504–518. doi:10.1109/TDSC.2014.2369034.

[66] J. M. Wu, J. Zhan, J. C. Lin, Ant colony system sanitization approach to hiding sensitive itemsets, IEEE Access 5 (2017) 10024–10039. doi:10.1109/ACCESS.2017.2702281.

[67] Q. Han, D. Lu, K. Zhang, X. Du, M. Guizani, Lclean: A plausible approach to individual trajectory data sanitization, IEEE Access 6 (2018) 30110–30116. doi:10.1109/ACCESS.2018.2833163.

[68] T. Barrabi, Why hackers love the cloud (2016) [cited 16.12.2016].
URL https://www.foxbusiness.com/features/why-hackers-love-the-cloud

[69] Y. Zhang, H. Huang, Y. Xiang, L. Y. Zhang, X. He, Harnessing the hybrid cloud for secure big image data service, IEEE Internet of Things Journal 4 (5) (2017) 1380–1388. doi:10.1109/JIOT.2017.2732357.

[70] L. Mearian, Internet of things data to top 1.6 zettabytes by 2022. (2016) [cited 25.12.2018].
URL ttps://campustechnology.com/articles/2015/04/15/internet-of-thingsdata-to-top-1-6-zettabytes-by-2020.aspx

[71] S. S. Yau, H. G. An, A. B. Buduru, An approach to data confidentiality protection in cloud environments, International Journal of Web Services Research 9 (3) (2012) 67–83. doi:10.4018/jwsr.20120701041.

[72] Z. He, Z. Cai, Y. Sun, Y. Li, X. Cheng, Customized privacy preserving for inherent data and latent data, Personal and Ubiquitous Computing 21 (1) (2017) 43–54. doi:10.1007/s00779-016-0972-2.
URL https://doi.org/10.1007/s00779-016-0972-2

[73] S. Sharma, K. Chen, A. Sheth, Toward practical privacy-preserving analytics for iot and cloud-based healthcare systems, IEEE Internet Computing 22 (2) (2018) 42–51. doi:10.1109/MIC.2018.112102519.

[74] M. Taneja, N. Jalodia, A. Davy, Distributed decomposed data analytics in fog enabled iot deployments, IEEE Access 7 (2019) 40969–40981. doi:10.1109/ACCESS.2019.2907808.

[75] J. Li, F. Guo, W. Jiang, Source localization and calibration using tdoa and fdoa measurements in the presence of sensor location uncertainty, Science China Information Sciences 57 (4) (2014) 1–12. doi:10.1007/s11432-013-4800-2.
URL https://doi.org/10.1007/s11432-013-4800-2

[76] A. Ferdowsi, U. Challita, W. Saad, Deep learning for reliable mobile edge analytics in intelligent transportation systems: An overview, IEEE Vehicular Technology Magazine 14 (1) (2019) 62–70. doi:10.1109/MVT.2018.2883777.

[77] D. Newman, Ttop 10 digital transformation trends for 2019, Forbes.

[78] R. H. Weber, Internet of things – new security and privacy challenges, Computer Law and Security Review 26 (1) (2010) 23 – 30. doi:https://doi.org/10.1016/j.clsr.2009.11.008.
URL http://www.sciencedirect.com/science/article/pii/S0267364909001939

[79] E. Pateromichelakis, F. Moggio, C. Mannweiler, P. Arnold, M. Shariat, M. Einhaus, Q. Wei, . Bulakci, A. De Domenico, End-to-end data analytics framework for 5g architecture, IEEE Access 7 (2019) 40295–40312. doi:10.1109/ACCESS.2019.2902984.

[80] B. Li, Y. Vorobeychik, M. Li, B. Malin, Scalable iterative classification for sanitizing large-scale datasets, IEEE Transactions on Knowledge and Data Engineering 29 (3) (2017) 698–711. doi:10.1109/TKDE.2016.2628180.

[81] A. Sharma, T. Goyal, E. S. Pilli, A. P. Mazumdar, M. C. Govil, R. C. Joshi, A secure hybrid cloud enabled architecture for internet of things, in: 2015 IEEE 2nd World Forum on Internet of Things (WF-IoT), 2015, pp. 274–279. doi:10.1109/WF-IoT.2015.7389065.

[82] B. Nguyen, N. Choi, M. Thottan, J. V. der Merwe, Simeca: Sdn-based iot mobile edge cloud architecture, in: 2017 IFIP/IEEE Symposium on Integrated Network and Service Management (IM), 2017, pp. 503–509. doi:10.23919/INM.2017.7987319.

[83] J. Delsing, Local cloud internet of things automation: Technology and business model features of distributed internet of things automation solutions, IEEE Industrial Electronics Magazine 11 (4) (2017) 8–21. doi:10.1109/MIE.2017.2759342.

[84] I. Lee, K. Lee, The internet of things (iot): Applications, investments, and challenges for enterprises, Business Horizons 58 (4) (2015) 431 – 440. doi:https://doi.org/10.1016/j.bushor.2015.03.008.
URL http://www.sciencedirect.com/science/article/pii/S0007681315000373

[85] Gartner, Gartner says worldwide public cloud services market to grow 18 percent in 2017 (2017) [cited 25.08.2019].
URL https://www.gartner.com/en/newsroom/press-releases/2017-02-22-gartner-says-worldwide-public-cloud-services-market-to-grow-18-percent-in-2017