Big Data Analytics in the Cloud for Future Smart Cities

Prof. Amresh Kumar[1] Ms. Sayali Baitule[2] Prof. Ashwini Yerlekar[3] Prof. Parag Dhawan[4]
[1,2]Department of Computer Technology [3]Department of Computer Science & Engineering [4]Department of Information Technology
[1,2,3,4]Rajiv Gandhi College of Engineering & Research, Nagpur, Maharashtra

Abstract— As time grows the need of the technology grows with increase in demand and support. ICT is grabbing the globe in every perspective of various technical as well as non-technical areas. Information and Communication Technology (ICT) [4] is the synonym for information technology (IT), enables users to access, store, transmit, and manipulate information. ICT helps in supporting organizational change, and played a primary role in the compliance of the various technological demands. ICT today is ever-present, with approximately 8 out of 10 Internet users owning a Smartphone, information and data are growing by very rapidly. This rapid growth, especially in developing countries, has led ICT to become a foundation of everyday life, in which life without some facet of technology turn into most of clerical job and schedule tasks dysfunctional. ICT involving Big Data (It includes Modeling, Services, Analytics, Economic & Business Analysis and Management Model, Government Management Models and Practices and Smart Planet Solutions.), Cloud and Security serves as a stage for business professional, academics, researchers, scientists, consultants and policy makers. Over the years governments all over the world have tried to take benefit of information and communication technology (ICT) to progress government operations and communication. Acceptance of e-government has increased in most cities and countries, but at the same time, the rate of successful acceptance and operation varies from city to city and country to country. ICT evolution and several inter-related institutional changes including government operations, public services delivery, citizen participation, policy and decision making, and governance reform. This Paper outlines the evolution of ICT and attempts to explore an idea on the role of big data analytics to accomplish the goal of making future smart cities and e-government management models and practices in holistic way..

Key words: ICT, Big Data, Big Data Analytics, Smart City, Cloud Computing

I. INTRODUCTION

A. ICT

According to the European Commission, the importance of ICTs lies less in the technology itself than in its ability to create greater access to information and communication in underserved populations. Many countries around the world have established organizations for the promotion of ICTs, because it is feared that unless less technologically advanced areas have a chance to catch up, the increasing technological advances in developed nations will only serve to exacerbate the already-existing economic gap between technological "have" and "have not" areas. Internationally, the United Nations actively promotes ICTs for Development (ICT4D) as a means of bridging the digital divide.

B. Big Data

Big data means really a big data, it is a collection of large datasets that cannot be processed using traditional computing techniques. From an evolutionary perspective, big data is not new. A major reason for creating data warehouses in the 1990s was to store large amounts of data. Back then, a terabyte was considered big data. Big data is not merely a data, rather it has become a complete subject, which involves various tools, techniques and frameworks.
Big Data includes huge volume, high velocity, and extensible variety of data.

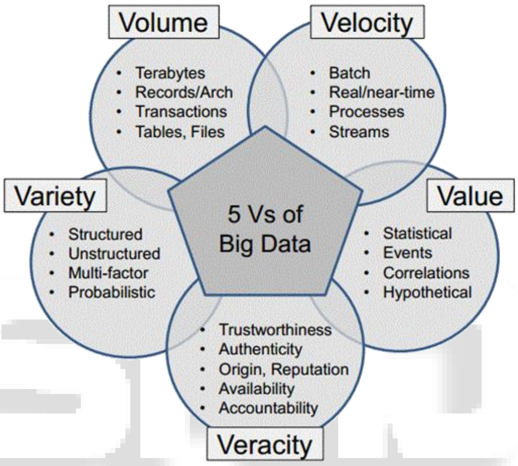

Fig. 1: 5V's of Big Data

The data in it will be of three types:

- Structured data: Relational data.
- Semi Structured data: XML data.
- Unstructured data: Word, PDF, Text, Media Logs.

We have entered the big data era. Organizations are capturing, storing, and analyzing data that has high volume, velocity, and variety and comes from a variety of new sources, including social media, machines, log files, video, text, image, RFID, and GPS.

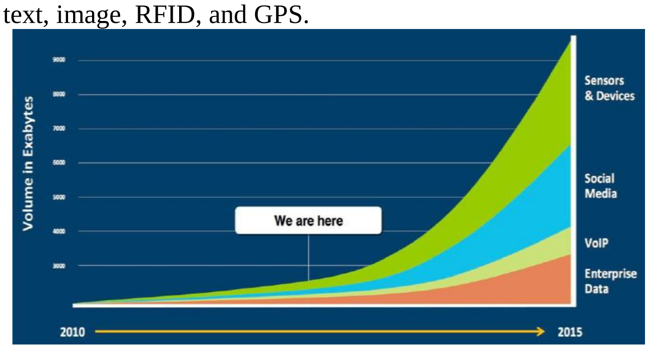

Fig. 2: The Exponential Growth of Big Data (Source: Palfreyman, 2013)

Big data is a term that is used to describe data that is high volume, high velocity, and/or high variety; requires new technologies and techniques to capture, store, and analyze it; and is used to enhance decision making, provide insight and discovery, and support and optimize processes.

C. Big Data Analytics

Big data analytics refers to the strategy of analyzing large volumes of data, or big data. This big data is gathered from a wide variety of sources, including social networks, videos, digital images, sensors, and sales transaction records. The aim in analyzing all this data is to uncover patterns and connections that might otherwise be invisible, and that might provide valuable insights about the users who created it. Through this insight, businesses may be able to gain an edge over their rivals and make superior business decisions.

Big data and analytics are hot topics in both the popular and business press. Big data and analytics are intertwined, but analytics is not new. Many analytic techniques, such as regression analysis, simulation, and machine learning, have been available for many years. Even the value in analyzing unstructured data such as e-mail and documents has been well understood. What is new is the coming together of advances in computer technology and software, new sources of data (e.g., social media), and business opportunity. This confluence has created the current interest and opportunities in big data analytics. It is even spawning a new area of practice and study called “data science” that encompasses the techniques, tools, technologies, and processes for making sense out of big data. It is helpful to recognize that the term analytics is not used consistently; it is used in at least three different yet related ways.

D. Cloud Computing

Cloud computing [3] also on-demand computing, is a kind of Internet-based computing that provides shared processing resources and data to computers and other devices on demand. This cloud model is composed of five essential characteristics, three service models, and four deployment models.

The NIST Definition of Cloud Computing lists five essential characteristics of Cloud Computing. It is reasonable to assume that missing any one of these essential characteristics means a service or computing capability cannot be considered as Cloud Computing.

1) On-demand Self-service. A consumer can unilaterally provision computing capabilities, such as server time and network storage, as needed automatically without requiring human interaction with each service provider.
2) Broad Network Access. Capabilities are available over the network and accessed through standard mechanisms that promote use by heterogeneous thin or thick client platforms (e.g., mobile phones, tablets, laptops, and workstations).
3) Resource Pooling. The provider's computing resources are pooled to serve multiple consumers using a multi-tenant model, with different physical and virtual resources dynamically assigned and reassigned according to consumer demand. There is a sense of location independence in that the customer generally has no control or knowledge over the exact location of the provided resources but may be able to specify location at a higher level of abstraction (e.g., country, state, or datacenter). Examples of resources include storage, processing, memory, and network bandwidth.
4) Rapid elasticity. Capabilities can be elastically provisioned and released, in some cases automatically, to scale rapidly outward and inward commensurate with demand. To the consumer, the capabilities available for provisioning often appear to be unlimited and can be appropriated in any quantity at any time.
5) Measured service. Cloud systems automatically control and optimize resource use by leveraging a metering capability at some level of abstraction appropriate to the type of service (e.g., storage, processing, bandwidth, and active user accounts). Typically this is done on a pay-per-use or charge-per-use basis. Resource usage can be monitored, controlled, and reported, providing transparency for both the provider and consumer of the utilized service.

So, Cloud Computing is measured, on-demand, elastic computing using pooled resources, usually on the Internet. Next, the NIST Definition of Cloud Computing list three service models a.k.a SPI model [1] [2](SPI: SaaS, PaaS, IaaS):

1) Software as a Service (SaaS). The capability provided to the consumer is to use the provider's applications running on a cloud infrastructure2. The applications are accessible from various client devices through either a thin client interface, such as a web browser (e.g., web-based email), or a program interface. The consumer does not manage or control the underlying cloud infrastructure including network, servers, operating systems, storage, or even individual application capabilities, with the possible exception of limited user-specific application configuration settings.More on Software as a Service (SaaS)
2) Platform as a Service (PaaS). The capability provided to the consumer is to deploy onto the cloud infrastructure consumer-created or acquired applications created using programming languages, libraries, services, and tools supported by the provider. The consumer does not manage or control the underlying cloud infrastructure including network, servers, operating systems, or storage, but has control over the deployed applications and possibly configuration settings for the application-hosting environment. More on Platform as a Service (PaaS).
3) Infrastructure as a Service (IaaS). The capability provided to the consumer is to provision processing, storage, networks, and other fundamental computing resources where the consumer is able to deploy and run arbitrary software, which can include operating systems and applications. The consumer does not manage or control the underlying cloud infrastructure but has control over operating systems, storage, and deployed applications; and possibly limited control of select networking components (e.g., host firewalls). More on Infrastructure as a Service (IaaS).

The increasing selection of services delivered over the Internet is sometimes referred to as XaaS.Finally, the NIST

Definition of Cloud Computing lists four deployment models:

1) Private cloud. The cloud infrastructure is provisioned for exclusive use by a single organization comprising multiple consumers (e.g., business units). It may be owned, managed, and operated by the organization, a third party, or some combination of them, and it may exist on or off premises. Community cloud. The cloud infrastructure is provisioned for exclusive use by a specific community of consumers from organizations that have shared concerns (e.g., mission, security requirements, policy, and compliance considerations). It may be owned, managed, and operated by one or more of the organizations in the community, a third party, or some combination of them, and it may exist on or off premises. More on Community Clouds.
2) Public cloud. The cloud infrastructure is provisioned for open use by the general public. It may be owned, managed, and operated by a business, academic, or government organization, or some combination of them. It exists on the premises of the cloud provider.
3) Hybrid cloud. The cloud infrastructure is a composition of two or more distinct cloud infrastructures (private, community, or public) that remain unique entities, but are bound together by standardized or proprietary technology that enables data and application portability (e.g., cloud bursting for load balancing between clouds).

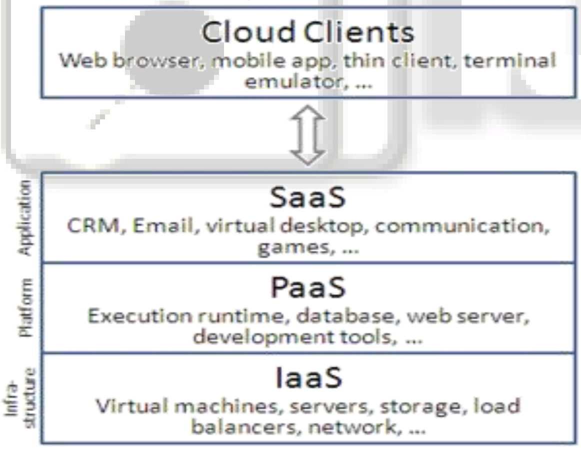

Fig. 3: Cloud computing lists four deployment models

E. Smart City

A smart city [4] is an urban development vision to integrate multiple information and communication technology (ICT) solutions in a secure fashion to manage a city's assets – the city's assets include, but not limited to, local departments information systems, schools, libraries, transportation systems, hospitals, power plants, water supply networks, waste management, law enforcement, and other community services. The goal of building a smart city is to improve quality of life by using technology to improve the efficiency of services and meet residents' needs.

Accordingly, the purpose of the Smart Cities Mission is to drive economic growth and improve the quality of life of people by enabling local area development and harnessing technology, especially technology that leads to Smart outcomes.

Application of Smart Solutions will enable cities to use technology, information and data to improve infrastructure and services. Comprehensive development in this way will improve quality of life, create employment and enhance incomes for all, especially the poor and the disadvantaged, leading to inclusive Cities.

II. BIG DATA ANALYTICS IN THE CLOUD

Businesses [6] have long used data analytics to help direct their strategy to maximize profits. Ideally data analytics helps eliminate much of the guesswork involved in trying to understand clients, instead systemically tracking data patterns to best construct business tactics and operations to minimize uncertainty. Not only does analytics determine what might attract new customers, often analytics recognizes existing patterns in data to help better serve existing customers, which is typically more cost effective than establishing new business. In an ever-changing business world subject to countless variants, analytics gives companies the edge in recognizing changing climates so they can take initiate appropriate action to stay competitive. Alongside analytics, cloud computing is also helping make business more effective and the consolidation of both clouds and analytics could help businesses store, interpret, and process their big data to better meet their clients' needs. Aside from its increased accessibility and utility, big data analysis on cloud drives also exports many IT demands, such as hosting and maintaining servers, to cloud service providers. Companies can spend less money on servers and instead focus on bolstering their staff and product. Thus, cloud drives help smaller companies get into the big data game, allowing start-ups to better compete with larger organizations in their industry.

Software-as-a-Service, or SaaS, is another popular function for cloud data analysis. SaaS lets users access software housed on a clouds remotely from any web browser, diminishing the need to use specific machines to accomplish a task. Typical uses for Saas included companies charging clients a membership or monthly fee to access software on their website so users only have access to the software as long as they pay their dues, never fully having the software on the hard drives of their own computers. While SaaS provides users flexibility concerning where they can access their applications, it also limiting if they do not have access to the internet or wish to work offline. Likely more software companies will adopt SaaS overtime as to further their profits and have complete control over their product. For users of the software this means they can either save money from having to buy software if it is only needed for a short time or cost them more money in long run as costs of subscription to the software adds up. Salesforce CRM, Google Apps, and DeskAway are all examples of SaaS.

III. ROLE OF BIG DATA ANALYTICS FOR FUTURE SMART CITIES

Approximately 50% of world's population lives in urban areas, a number which is expected to increase to nearly 60% by 2030. High levels of urbanization are even more evident in Europe where today over 70% of Europeans live in urban areas, with projections that this will increase to nearly 80% by 2030. A continuous increase in urban population strains the limited resources of a city, affects its resilience to the increasing demands on resources and urban governance faces ever increasing challenges. Furthermore, sustainable urban development, economic growth and management of natural resources such as energy and water require better planning and collaborative decision making at the local level. In this regard, the innovation in ICT can provide integrated information intelligence for better urban management and governance, sustainable socioeconomic growth and policy development using participatory processes.

Smart cities use a variety of ICT solutions to deal with real life urban challenges. Some of these challenges include environmental sustainability, socioeconomic innovation, participatory governance, better public services, planning and collaborative decision-making. In addition to creating a sustainable futuristic smart infrastructure, overcoming these challenges can empower the citizens in terms of having a personal stake in the well-being and betterment of their civic life. Consequently, city administrations can get new information and knowledge that is hidden in large-scale data to provide better urban governance and management by applying these ICT solutions. Such ICT enabled solutions thus enable efficient transport planning, better water management, improved waste management, new energy efficiency strategies, new constructions and structural methods for health of buildings and effective environment and risk management policies for the citizens. Moreover, other important aspects of the urban life such as public security, air quality and pollution, public health, urban sprawl and bio-diversity loss can also benefit from these ICT solutions. ICT as prime enabler for smart cities transforms application specific data into useful information and knowledge that can help in city planning and decision-making. From the ICT perspective, the possibility of realisation of smart cities is being enabled by smarter hardware and software e.g. IoTs i.e. RFIDs, smart phones, sensor nets, smart household appliances, and capacity to manage and process large scale data using cloud computing without compromising data security and citizens privacy. With the passage of time, the volume of data generated from these IoTs is bound to increase exponentially and classified as Big data. In addition, cities already possess land use, transport, census and environmental monitoring data which is collected from various local, often not interconnected sources and used by application specific systems but is rarely used as collective source of information (i.e. system of systems) for urban governance and planning decisions. Many local governments are making such data available for public use as “open data”. Managing such large amount of data and analyzing for various applications e.g. future city models, visualization, simulations, provision of quality public services and information to citizens and decision making becomes challenging without developing and applying appropriate tools and techniques.

In the above context, recent emergence of Cloud computing promises solutions to such challenges by facilitating big data storage and delivering the capacity to process, visualize and analyze city data for information and knowledge generation. Such a solution can also facilitate the decision makers in meeting the QoS requirements by providing an integrated information processing and analytic infrastructure for variety of smart cities applications to support decision-making for urban governance.

Figure below [5] depicts our view of the main thematic pillars of smart cities: smart people, smart economy, smart environment, smart governance and smart mobility which contribute towards the sustainability of resources and resilience against increasing urban demands. The main motive towards developing such a view is to consider a holistic approach for smart cities by providing data acquisition, integration, processing and analysis mechanisms to synthesize the needed information that can help in enhancing resilience and sustainability of a city. Managing data for these thematic domains in a Cloud environment provides the opportunity to integrate data acquired from various sources, process and analyse it in acceptable time-frames. However, it is not straightforward to adopt cloud computing to deal with smart city applications due to a number of challenges and requirements. Our aim here is to discuss a perspective on how these challenges can be addressed in part by using ICT tools and software services to intelligently manage and analyse the complex big data of smart cities, by incorporating a suitable Cloud architecture.

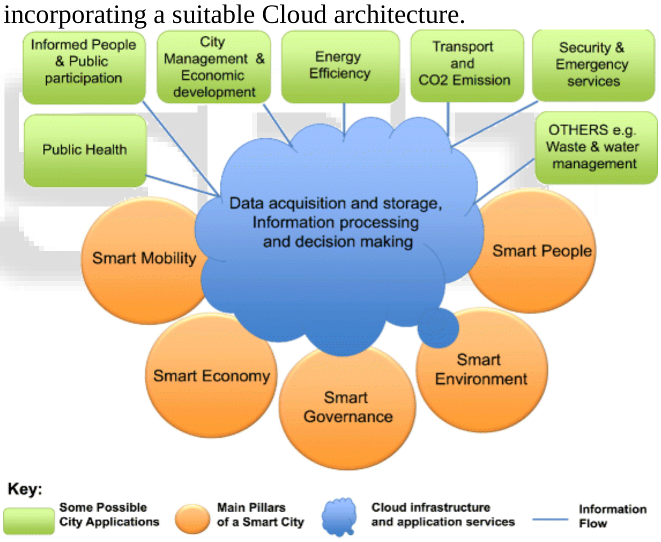

Fig. 4: Cross-thematic data management and analysis for variety of smart city applications in Cloud environment. [5]

Big Data Discovery[7][8] is the next big trend in analytics: It’s the logical combination of three of the hottest trends of the last few years in analytics: Big Data, Data Discovery, and Data Science [5].

Each of these areas has seen explosive growth, but there are clear upsides and downsides to each. For example, Data Discovery excels in ease of use, but allows only limited depth of exploration, while Data Science provides powerful analysis but is slow, complex, and difficult to implement.

Since the disadvantages of the three technologies map to nicely to the advantages of the others, they are now starting to blend, and Gartner believes Big Data Discovery will be a distinct new market category by 2017.

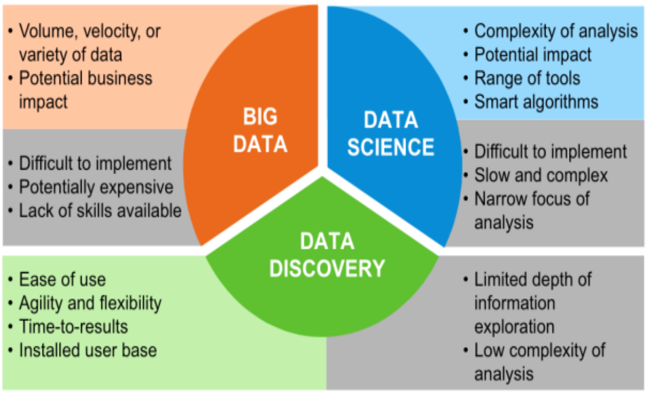

Fig. 5: Data Discovery is the combination of Big Data, Data Science, and Data Discovery [5].

IV. Challenges Facing Big Data and Analytics [8]

A. *Meeting the Need For Speed*

In today's hypercompetitive business environment, companies not only have to find and analyze the relevant data they need, they must find it quickly. Visualization helps organizations perform analyses and make decisions much more rapidly, but the challenge is going through the sheer volumes of data and accessing the level of detail needed, all at a high speed. The challenge only grows as the degree of granularity increases. One possible solution is hardware.

B. *Understanding The Data*

It takes a lot of understanding to get data in the right shape so that you can use visualization as part of data analysis. Without some sort of context, visualization tools are likely to be of less value to the user.
One solution to this challenge is to have the proper domain expertise in place. Make sure the people analyzing the data have a deep understanding of where the data comes from, what audience will be consuming the data and how that audience will interpret the information.

C. *Addressing Data Quality*

Even if you can find and analyze data quickly and put it in the proper context for the audience that will be consuming the information, the value of data for decision-making purposes will be jeopardized if the data is not accurate or timely. This is a challenge with any data analysis, but when considering the volumes of information involved in big data projects, it becomes even more pronounced. Again, data visualization will only prove to be a valuable tool if the data quality is assured. To address this issue, companies need to have a data governance or information management process in place to ensure the data is clean. It's always best to have a pro-active method to address data quality issues so problems won't arise later.

D. *Displaying Meaningful Results*

Plotting points on a graph for analysis becomes difficult when dealing with extremely large amounts of information or a variety of categories of information. One way to resolve this is to cluster data into a higher-level view where smaller groups of data become visible. By grouping the data together, or "binning," you can more effectively visualize the data.

V. Discussion and Conclusion

In this paper we discussed the cloud based big data analytics for smart future cities. Several considerations need to be carefully planned. Due to multidisciplinary nature of smart city application domains, engagement with domain experts is needed to identify basic relationships and dependencies between different data elements. Big Data is an essential component that is driving the Smart Cities movement, along with more general advances in technology.

This is informing new forms of consumption with citizens and the demand for services that are underpinned by smarter systems. The social contract for taking care of citizens remains for cities that prompt its managers to consider and invest in appropriate processes and systems to support this mandate. In this, certain best practices have emerged and others are indeed still to emerge, as this is a dynamic environment. Cities that make use of the current state-of-the-art approaches, which can be benchmarked against international standards, are positioned for future success and may have the opportunity for advancing society forward to a smarter future.

References

[1] http://searchcloudcomputing.techtarget.com/definition/SPI-model
[2] http://www.paranet.com/blog/bid/128267/The-three-types-of-Cloud-Computing-Service-Models
[3] http://www.service-architecture.com/articles/cloud-computing/cloud_computing_definition.html
[4] http://smartcities.gov.in/writereaddata/What%20is%20Smart%20City.pdf
[5] http://journalofcloudcomputing.springeropen.com/articles/10.1186/s13677-015-0026-8
[6] https://www2.deloitte.com/content/dam/Deloitte/fpc/Documents/services/systemes-dinformation-et-technologie/deloitte_smart-cities-big-data_en_0115.pdf
[7] http://timoelliott.com/blog/2015/03/what-is-big-data-discovery.html
[8] https://www.sas.com/resources/asset/five-big-data-challenges-article.pdf
[9] http://www.oracle.com/us/products/oracle-big-data-executive-brief-2415577.pdf.