Application of Business Analytics on Big Data Scenarios to obtain Actionable Insights in Systems or Services in Mega-Cities: Multiple use cases from the Living Lab at the Smart Cities Innovation Center at Guadalajara University

Manuel Avalos, *Ignacio Silva, Victor M. Larios, IEEE*

IEEE GDL Physical Infrastructure Working Group

Abstract—Cities, organizations and intelligence agencies from a variety of industries and sectors face the same challenges regarding Big Data: the need to uncover actionable insights with limited resources and with critical speed. However, with the right intelligent solution, overwhelming data will not be a challenge. Instead, such data become a resource for smarter, more effective decision making. This research will focus on the application of business analytics to attain actionable insights from data deployed in a Hybrid Cloud and providing that "curated" data to the decision makers in cities, mega-cities, companies, and city service providers. This paper intends to be a best-practice guideline for cities interested in implementing Big Data use cases to solve some of their main problems.

Index Terms— M.8.8 Cloud computing, M.10.0.d Service cloud and ecosystems, M.11.4.d Business intelligence, M.12.4.a Big Data technology, M.12.3.d Data transformation (http://www.computer.org/mc/keywords/keywords.htm)

1 INTRODUCTION

Governments are confronting serious challenges that affect their economies and their ability to deliver core services to their citizens (3). They are faced with the harsh realities of swelling city populations that demand more services and have aging infrastructure, declining budgets, and increasing threats. To help address these challenges, governments are looking at cloud computing as a way to acquire new capabilities more quickly. With less investment upfront, this type of solution enables cities to become "smarter" by intelligently connecting city events and data in their jurisdictions with a single integrated solution (2).

IBM´s view on a smarter planet has 3 key elements:

1. Instrumented: Sensors everywhere getting data from multiple sources
2. Interconnected: Telecommunications, data gathering
3. Intelligence: Getting insight from the collected data for better outcomes and decisions

For more than 6 years, companies, cities, and communities around the world have been helping build a Smarter Planet. There have been significant advances as leaders have begun using the vast supply of data to transform their enterprises and institutions through Big Data and analytics, mobile technology, social business, and the cloud.

Data: the new natural resource (1)

Executives have traditionally regarded experience and intuition to be the keys to formulating strategy and assessing risk.

That type of thinking may have worked in an earlier time of information scarcity but does not work in the time of Big Data.

When every company, every city, every country, and every individual is increasingly interconnected with millions of others, the cost of a poor decision can be devastating. However, analytics is increasingly helping business and government leaders look beyond their own biases to discern real patterns and anticipate events (1).

Social: the new production line

"Empower people. Understand people. Trust people."

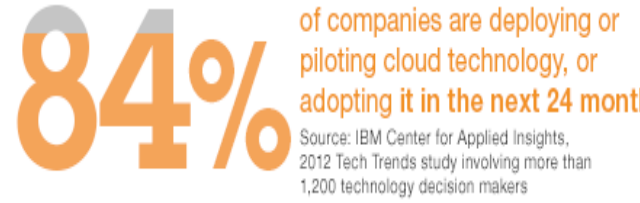
In today's knowledge economy, the exchange of ideas has become the new means of production. The advent of social and mobile technology is shifting employers' competitive edge from hiring workers who gather knowledge to hiring workers who communicate knowledge.

By 2020, because of the pervasive adoption of social technologies, new systems of people-centric engagement will be mainstream. Successful enterprises are now able to tap into shared insights, collective knowledge, and expertise at the individual level to empower more meaningful engagement with both employees and customers.

Cloud: the new growth engine

Today's cloud is about more than expanding reach. It is about expanding possibility. Once, the cloud was simply the tool you used to access your files from anywhere and to make your operations more efficient. Today, the cloud is an incubator of ideas, where one can continually test, adjust, and deploy innovations faster than ever before.

As explained in the "Demystifying The Cloud" article from Data House (6), from Phase I: Mainframe Computing to Phase II: Distributed Computing, the current Phase III: Cloud computing offers the central management and coordination of mainframe computing with the affordability and scalability of distributed computing. The cloud is therefore an ideal solution for both large and small companies. IBM Bluemix has an extensive catalog of services from IBM, third party services, and community services to extend the functionality of client applications. This PaaS solution from IBM also helps clients deploy and manage hybrid applications. Bluemix provides a single development and management experience across any combination of public, dedicated, and local Bluemix instances. Users choose where their apps, data, and other services live without compromising the speed or economics of the cloud using open-source technologies, such as "Docker-IBM Containers," which offer portable and consistent delivery of applications without having to manage an OS. (7)



With the convergence of mobile networks, social media, and analytics, organizations are discovering entirely new ways to engage with their customers, employees, and partners. Moreover, with the cloud, this engagement occurs quickly, iteratively, and with very little risk. One example of this could be IBM Bluemix (Platform As A Service), which allows clients to start with trial versions and easily build their applications (their way), using a combination of the most prominent open-source computing technologies to power their applications. Once developed, Bluemix services handle the rest. (8)

Mobile: the new office space

The term "mobile" for business has evolved quickly and no longer means a web site optimized for a smart phone. It means empowering employees to be effective no matter where they are while protecting the infrastructure. It means offering customers a way to do business with the company anytime and anywhere. It also means launching new business models based on the computer that most people carry in their pocket (1).

2 PERSPECTIVES

2.1 Review Stage

Cloud: How city managers can lower costs and improve services

Globally, city and municipal government leaders recognize the importance of cloud computing for tying budgets directly to service consumption and lowering upfront capital costs compared to a traditional deployment. According to a survey by the Public Technology Institute (PTI), 45 percent of local governments are using some form of cloud computing and another 19 percent plan to do so by the end of 2011. PTI report-

ed that resource savings (for example, staff time and maintenance and support costs) was a common reason for local governments to turn to the cloud. A former United States Chief Information Officer, in his Federal Cloud Computing Strategy, estimated that US federal datacenter infrastructure costs could drop by 30 percent through the use of a cloud-computing model, amounting to approximately $7.2 billion in total savings. **Reducing capital costs and operating expenses with the cloud.**

To help governments realize the benefits of cloud computing, IBM is introducing Smarter City Solutions on Cloud, a new software-as-a-service (SaaS) offering. The solution is available on the IBM SmartCloud Enterprise, a public cloud platform that includes hardware, network, and storage. It is designed for cities of all sizes to manage their operations and facilitate collaboration between multiple agencies. By providing cities with a comprehensive "pay-as-you-go" solution that includes application software, infrastructure, networking, systems software, middleware, and maintenance, it lowers the barrier of entry while enabling cities to deliver excellent service to their citizens.

The benefits of the cloud are compelling. Early adopters of cloud computing, including IBM's own IT organization and IBM customers, have already realized significant savings. These savings include reduced IT labor costs by up to 50 percent in configuration, operations, management, and monitoring, as well as lowering IT support costs by up to 40 percent.

Additional financial benefits for running applications in a cloud model include lower upfront capital costs than with a traditional application deployment. Actual savings will depend on factors such as capital budget restrictions, size and complexity of integrating applications, existing operating expenses, IT infrastructure capacity, and the skills of the existing IT staff.

Operations benefits on the Cloud

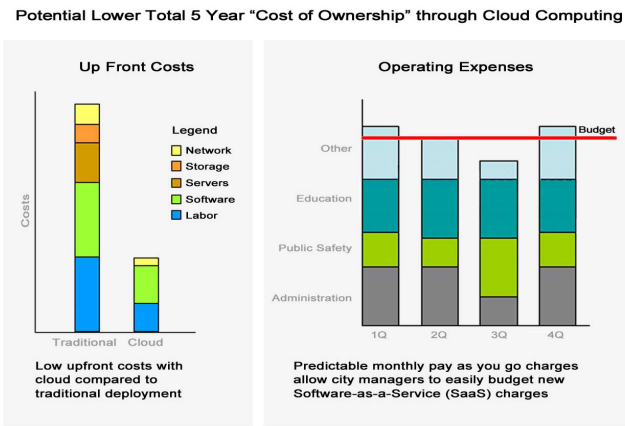


Figure 1: Potential lower total cost of ownership compared to traditional deployment and predictable operating expenses may align better with the cash flow and operating requirements for city governments (Comparative models are based on IBM internal projections. Actual costs and implementation times might vary by customer).

Defining a S*marter City*

IBM defines a Smarter City in terms of three key characteristics: instrumented, interconnected, and intelligent. These characteristics suggest that a city must be able to sense and respond to its environment. A cities government can observe its surroundings with sensors and other data-gathering mechanisms (instrumented), communicate and share information gathered with other government entities (interconnected), and use the data to make optimal service choices (intelligent). To make truly intelligent decisions requires data to be collected and analyzed from a diverse group of regional and local agencies.

With the right instruments, each of the service departments (**Figure 2**) has the capability to provide cities the opportunity to gather a tremendous amount of detail about the city's physical environment. Details from diverse sources such as traffic flow, electrical usage, water quality, and citizen emergencies are then analyzed to make predictions about resource needs and intelligent operations

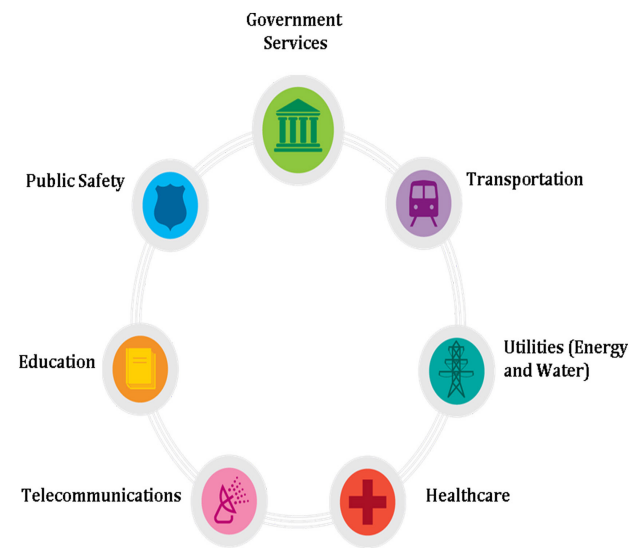


Figure 2: Departments and services in a Smart City

At this point, it is important to understand the definitions of a smarter city by key players in the ecosystem, such as IBM, Cisco, and IEEE**.** Although Cisco is considering the Internet of everything to be the base of the smart city (4), IEEE has a more holistic point of view, connecting people, government, and industry, with the technology as the vehicle of innovation for a better quality of life and environmental sustainability (5).

Finally, IBM plans to apply different service models to city services, such as private, hybrid or public, as shown in the figure 3 below.
In addition, the Smart Cities Innovation Center located at the Business School (CUCEA) campus in the Universidad de Guadalajara, and the IEEE Smart Cities Initiative in Guadalajara are using KPI to manage city operations based on Cohen´s wheel (5).

Figure 3: IBM's view for Smarter Cities on the Cloud

2.2 Big Data Scenarios

Remember what life was like before big data? The term 'Big Data' has become so prevalent in the business lexicon that it can be difficult to remember that Big Data are relatively recent phenomena. Some may have viewed Big Data as a fad, but data generated by people, processes, and machines are only continuing to grow. Big Data are here to stay.

Make no mistake, data are an asset but not when the user is drowning in them. In the information age, one of our greatest resources can also be our biggest downfall if our organizations do not know how to leverage data properly. Therefore, what can be done with the data?

Consider these actual scenarios:

- The healthcare industry spends roughly USD250 billion on healthcare fraud per year. By 2016, this could grow to more than USD400 billion a year. The US healthcare sector could create more than USD300 billion in value every year using Big Data creatively and effectively to drive better efficiency and quality.

- One rogue trader at a leading global financial services firm created USD2 billion worth of losses, almost bankrupting the company. Financial institutions now have a lot more data readily available to help them prevent both external fraud (involving customers, account holders, or policyholders) and internal, employee-related incidents.

- In Europe, governments could save more than EUR100 billion (USD149 billion) in operational efficiency improvements alone using Big Data, not including using Big Data to reduce fraud, errors and boost the collection of tax revenues.

- Retailers miss USD93 billion in sales each year because they do not have the right products in stock to meet customer demand. A retailer using Big Data could increase its operating margin by more than 60 percent.

- Six billion global subscribers in the telecommunications industry, which is growing at double-digit rates each year, are demanding unique, personalized, and often location-based offerings that match their individual lifestyles.

With such high-stakes costs and opportunities, the market is primed for Big Data solutions. In a recent study conducted by the IBM Institute for Business Value in collaboration with the "Saïd Business School at the University of Oxford", respondents were asked to describe the level of Big Data activities in their organizations today. The results suggest four main stages of Big Data adoption and progression along a continuum: Educate, Explore, Engage and Execute.

Although only 6 percent of organizations are already executing Big Data initiatives, approximately one-quarter of all organizations are piloting initiatives, another half of all organizations are developing a strategy and will be looking to purchase soon, and the final quarter of all organizations are in an information-gathering phase.

If companies are not working on Big Data strategies, then their competitors probably are. The difficulty is determining how and where to get started.

2.3 Big Data Use Cases

Because much of the Big Data activity in the market in the past has focused on learning about Big Data technologies, vendors have not made a concerted effort to help organizations understand which problems Big Data can address. IBM has been the exception.

By conducting surveys, studying analyst findings, talking with more than 300 customers and prospective customers, and implementing hundreds of Big Data solutions, IBM has identified the top five high-value use cases that can serve as a first step into Big Data (2):

1. **Big Data exploration:** Find, visualize, and understand Big Data to improve decision-making
2. **Enhanced 360-degree view of the customer:** Extend existing customer views by incorporating additional internal and external information sources
3. **Security/intelligence extension:** Reduce risk, detect fraud, and monitor cyber-security in real time

4. **Operations analysis:** Analyze a variety of machine data for improved business results and operational efficiency
5. **Data Warehouse Modernization:** Integrate Big Data and traditional data warehouse capabilities to gain new business insights while optimizing the existing warehouse infrastructure

There has already been a significant amount of work done in the area of Big Data applications for Business Intelligence and analytics, specifically, in the e-Commerce area, such as the "Muli-agent architecture for real-time Big Data processing" (10).

From the above cases, the authors are working closely with the UDG "Smarter Cities Innovation Center" (9) to deploy Hybrid Cloud environment capabilities. The authors will partner with other researchers connected to the CUCEA PhD program, who will collect data from sensors, reuse datasets captured from previous projects, use social collaboration data, etc. so we can execute a few use cases with Hybrid Cloud capabilities and Development platforms such as IBM BlueMix (8) and the potential use multi-agents platforms. **Exploration and Security-Intelligence as exposed in the Scope of this project**.

3 PROBLEMS, LIMITATIONS, DELIMITATIONS AND SCOPE

This study will mainly focus on the implementation of the different Big Data use cases outlined in the previous section on Hybrid Cloud deployment models. The idea is to implement a pilot program (Proof of Concepts) at the Universidad de Guadalajara CUCEA campus and then in a multi-campus environment within the Universidad de Guadalajara system. Because the CUCEA campus is now certified as a Living LAB for Smarter Cities Solutions, it is recognized officially by the State Government of Jalisco at the Science and Innovation ministry level. This CUCEA Living LAB will be certified every year and is currently part of a network of five Living LABs at the state level in Jalisco, and the government has pledged its support because of the potential value the LABs can bring to the State of Jalisco. The project is capable of succeeding and demonstrating some easy wins that can be implemented at the state level. Furthermore, the project and the Living LAB have the potential to be key success factors in the implementation of Smarter Cities solutions in the area.

Problems: Traffic, city services, lack of citizen engagement, siloed solutions, security threads, etc.

Limitations: Lack of infrastructure (Internet of Things "IoT" - sensors, etc.), isolated systems, lack of centrally controlled operation, lack of political willingness to collaborate across different institutions, and the infrastructure deployed in the cities in the "Zona Metropolitana de Guadalajara (ZMG)" that will present some technical challenges for potential integration and synergies.

Delimitations: Define small prototype projects (Proof of Concepts) that can later scale to higher levels by taking advantage of Hybrid Cloud Deployment models and Big Data use cases.

Scope: Implement one or, if possible, two Big Data use cases using Hybrid Cloud implementations at the "Centro de Innovacion en Ciudades Inteligentes" facility at CUCEA campus in the Universidad de Guadalajara (UDG) as part of the LIVING LAB Concept With a working prototype or Proof Of Concept (POC), the value added by expanding this POC to the UDG at-large and later to the cities surrounding the ZMG (Guadalajara, Zapopan, Tlaquepaue, Tonala, and the Tlajomulco municipalities) can be demonstrated.

One potential area for exploration expressed by politicians and citizens of the ZMG is security. This can be explored using Image Analytics and Social Networks data to gather information and insights into criminal events and attacks to the city and their citizens.

4 HYPOTHESIS

The implementation of some Big Data use cases, such as Exploration and Security-Intelligence, provides a method to demonstrate that using Hybrid Cloud technologies and gaining insights from a collected dataset from multiple sources, we can become smarter by completing a Prototype or Proof of Concept (POC) at UDG "Smarter Cities Innovation Center" (9). This can then be extended to the UDG at-large, where some of the best practices from the ZMG municipalities can be implemented. The goal is to start small with some POCs and pilots that will generate some **Wins/Successes, which will first garner support and confidence from the key stakeholders**, followed by scaling and growth.

5 OBJECTIVES

Main objective:

- Implement and Deploy Big Data use cases at the Universidad de Guadalajara (Pilot phase): the target use cases are Exploration and Security-Intelligence.
- Implement and Deploy one or two Big Data use cases at the UDG "Smarter Cities Innovation Center", then scale to the UDG at-large as a potential phase II.
- Finally, at Phase III, after achieving some quick Win/Successes, the community and stakeholders will gain confidence, followed by

a push to deploy the innovations in other municipalities of the ZMG or the ZMG at-large.

Specific objectives:

- Implement a small Proof of Concept (PoC) "1. Big Data exploration: use case" using the Hybrid Cloud model.
- Develop a small prototype at CUCEA "Smarter Cities Innovation Center".
- Embrace Open Data frameworks as a way to share data among city systems, services and citizens.
- Develop a best-practice guideline to move from PoC to deployment projects that can have an economic and social impact on cities and their citizens.

6 METHODOLOGY

- **State of the art.** Today, there are multiple analytical tools that can extract insight data from Big Data repositories; however, many systems are isolated and are not integrated in a holistic way; therefore, Open Data are not widely adopted.
- **Resources.** Existing Open Data frameworks and methodologies will be used, as will Hybrid Cloud solutions, to pilot some of the Big Data use cases focused on Exploration and Security-Intelligence use cases. The analytical tools used may be Open or industry proprietary. PaaS IBM Bluemix with Containers (Dockers Open Source technology) will be used for data portability.
- **Proposed model.** Divide the project into phases, where Phase 1 might be a Prototyping model on a small scale at the UDG Smarter Cities Innovation Center on the CUCEA Campus, where Phase II can subsequently be expanded to UDG at-large with a multi-campus approach. During Phase III, the model will be scaled and grown to benefit a city or megacity, taking advantage of the fact that the project location is a recognized LIVING LAB according to the State Government of Jalisco.
- **Deployment.** The Big Data use case for CUCEA UDG, as part of the "Smarter Cities Innovation Center," will be deployed in a Cloud Hybrid model and Exploration and Security-intelligence Big Data use cases.
- **Experimentation.** First, much experimentation will be conducted internally in a private domain and externally using Public Cloud services via IBM Bluemix PaaS and the wide and increasing service portfolio, such as a Cognitive System like Watson and Containers with Dockers open source code, ensuring that all aspects of application performance and security are covered during the development of the project
- **Analysis of results.** At each phase of the project, the results will be analyzed using **Vs** scope of work. Additionally, expected results, lessons learned, and adjustments for subsequent phases of the project will be investigated.

7 TIMELINE OF ACTIVITIES

7.1 Timeline and Activities

- 1H15 Create a private Cloud
- 1H15 Subscribe to a Public Instance
- 1H15 Develop a small Big Data use case prototype
- 3Q15 Test application in a Hybrid Cloud environment
- 4Q15 Deploy the application for small prototyping adoption (~1000 users+)
- 1H16 Deploy at Higher Scale (~10-20 K users)
- 2H16 Deploy at City level (100 K-1 Million users)

Acknowledgements

We appreciate the support of IBM, the IEEE Smart Cities initiative, the University of Guadalajara PhD in IT program and the Living Lab in the Smart Cities Innovation Center at CUCEA UDG.

REFERENCES

[1] Rometty, G. (2008, January 1). What's new on a Smarter Planet? Retrieved from http://www.ibm.com/smarterplanet/us/en/overview/ideas/
[2]IBM Software for Big Data ftp://ftp.software.ibm.com/software/pdf/se/8octBig_data_top_five.pdf
[3] The Smarter Cities Challenge (n.d.). Retrieved from http://smartercitieschallenge.org/
[4] The Internet of Everything from CISCO http://www.cisco.com/web/strategy/docs/gov/everything-for-cities.pdf
[5] The IEEE Smart Cities Initiative http://smartcities.ieee.org/
[6] Assets for Cloud Computing Architectures http://www.datahouse.com/assets/files/Cloud-1.0.pdf
[7] The IBM Dockers Platform for Public Clouds http://www.docker.com/
[8] https://console.ng.bluemix.net/

[9] The first Smart Cities Innovation Center at the University of Guadlajara
http://www.cucea.udg.mx/?q=noticia/todo-cucea-ser-laboratorio-de-ciudad-inteligente
[10]http://ieeexplore.ieee.org/xpl/articleDetails.jsp?reload=true&arnumber=6928203
[11] B. P. Rimal, E. Choi, and I. Lumb, "A Taxonomy and Survey of Cloud Computing Systems," *2009 Fifth International Joint Conference on INC, IMS and IDC*, pp. 44–51
[12] H. Schaffers, N. Komninos, M. Pallot, B. Trousse, M. Nilsson, and A. Oliveira, "Smart Cities and the Future Internet: Towards Cooperation Frameworks for Open Innovation," *Future Internet Assembly*, vol. 6656, no. 31, pp. 431–446, 2011

Authors' Biographies

Manuel Avalos is currently a PhD student at Universidad de Guadalajara and is a Global Product Manager within IBM Systems-Storage division, responsible for several Cloud Storage products-solutions. Manuel holds a Masters degree in Information Technology from ITESM (Monterrey Tech) and several certificates from Guadalajaran institutions in areas such as marketing, finance, and programming languages. Manuel is also an IBM Inventor and holds a US Patent **US 8010615 B2 "Instant messaging multilingual configuration"** and several patent publications at the IP.COM site. Manuel is a recognized Subject Matter Expert (SME) in areas of CAMS (**Cloud**, Analytics, Mobile & Social) at IBM and within the IT Ecosystem and is an active IBM Ambassador with multiple Universities in Jalisco.

Ignacio Silva-Lepe is a Research Staff Member at IBM Research. His areas of interest include (1) Component Software, designing and building application server infrastructure for distributed components; (2) Distributed Messaging Systems, providing reliable messaging infrastructure, and integrating messaging and web services technologies; (3) Advanced Enterprise Middleware, designing and prototyping techniques for hosted SOA environments and decentralized SOA federations; and (4) PaaS Research, designing and building infrastructure for on-boarding and instantiating platform as a service offerings onto a compute cloud.
Before joining IBM Research, Ignacio was a Member of the Technical Staff at Texas Instruments' Corporate Research and Development, subsequently acquired by Raytheon. Prior to that, he was a Research Assistant at Northeastern University, where he earned a PhD in Computer Science. He received his undergraduate degree in Computer Systems Engineering from Universidad ITESO, in Guadalajara, Mexico.

Victor M. Larios has received his PhD and a DEA (French version of a MS program) in Computer Science at the Technological University of Compiègne, France and a BA in Electronics Engineering at the ITESO University in Guadalajara, Mexico. He works at the University of Guadalajara (UDG), holding a Full Professor-Researcher position at the Department of Information Systems, and is the director of the Smart Cities Innovation Center at CUCEA UDG Campus. Moreover, Dr. Victor M. Larios founded the UDG PhD in Information Technologies in 2007 and has been leading projects in Guadalajara within the academy, government, and high-tech industry (IBM, Intel, and HP), focusing his research on distributed systems, parallel computing, data analytics and visualization, serious games, and Smart Cities. In 2013, Dr. Victor M. Larios was a consultant for and elected as the Director of Science and Technology for the Guadalajara Ciudad Creativa Digital A.C. project. Since July 2013, Dr. Victor M. Larios has volunteered as the Guadalajara City Local Leader for the IEEE Smart Cities Initiative. In addition, Dr. Victor M. Larios is an IEEE Senior member with a valued membership of 22 years and is the current secretary of the IEEE Computer Chapter at the Guadalajara Section.