CHAPTER 9

The Influence of Big Data and IoT on Smart Cities

Ioannis Karamitsos, Charalampos Manifavas, and Muhieddin Amer

INTRODUCTION

Smart city initiatives produce a large amount of data and information, collected through various sources. With the introduction of new technologies such as the Internet of Things (IoT) and Big Data, collected data can be accumulated, analyzed, and elaborated on to ultimately obtain valuable insight. The combination of Big Data and IoT is a research area that presents interesting challenges for the future of smart cities. These new challenges are related to the integration of business and technical streams into the smart city ecosystem. In this chapter, we present an overview of the Big Data and IoT technologies and show how they interact while converging. Several smart cities applications are analyzed within the context of the IoT and Big Data. Moreover, a future integration framework is proposed and research challenges are identified. This study can serve as a guideline for industries or stakeholders who engage in the development of smart cities initiatives in the context of Big Data and IoT. Many real estate companies have developed innovative smart city visions based on a portfolio of products and services for

---

I. Karamitsos (✉) • C. Manifavas • M. Amer
Rochester Institute of Technology – Dubai, Dubai, United Arab Emirates
e-mail: ixkcad1@rit.edu; cxmcad@rit.edu; mxaeee@rit.edu

W. A. Samad, E. Azar (eds.), *Smart Cities in the Gulf*,
https://doi.org/10.1007/978-981-13-2011-8_9

the real estate market. Such visions were derived after several years of innovation and on-the-field experience both in mature and developing economies. Integrated facility management platforms, real estate infrastructure convergence to IP, machine to machine (M2M) technologies, IoT, real-time business, and e-health are only a few building blocks of a solution that delivers enormous benefits to the real estate industry.

The new concept of connected devices over the existing ICT network has been developed with the appearance of smart devices. The past few years have seen a tremendous growth of devices connected to the network, and thus expanding the boundaries of conventional networks. This new configuration of connectivity has introduced the IoT technology. IoT is a technology which appeared first in 1999, as Ashton (2009) described it as an Internet evolution at MIT's AutoID lab. More specifically, IoT can be regarded as a concept under which a plethora of things/objects can be connected via wireless and wired connections over the internet. The IoT concept has become fairly matured with the attention of multiple research groups providing productive applications like smart home, smart city, smart transportation, and so forth. In the smart city context, the aim is to provide smart services to improve the quality of life through connected adaptive devices. Smart devices produced a large amount of data on a real-time basis for collection and elaboration from the aggregation devices. Due to the high volume of collected data, it is very difficult with the existing mechanisms to satisfy the real-time processing demand. Hence, the engagement and collaboration with the Big Data analytics become mandatory as a first step for a smarter city.

In this chapter, IoT and Big Data analytics are integrated with the smart city architecture for the deployment of smart cities. The proposed architecture is capable of handling different types of data formats structured, unstructured, and real-time decision-making. However, the data orchestration layer is the most influential component of the total framework for the realization of a smart city. In this study a hub approach and Hadoop are chosen as the storage and processing medium for the heterogeneous data.

Related Work

The concept of the rapid development of a smart city has attracted many researchers and architects toward studies on efficient framework and standard design. Standardizing the smart city models can provide various benefits from researchers in different standalone frameworks. Many research projects have covered different approaches ranging from abstract concepts to a complete set of services.

Nowadays, many researchers are working to present various solutions based on IoT technologies for smart cities. Similarly, various test bed simulations have been proposed to overcome the challenges. A full set of smart city services on various modules has been proposed by Sanchez et al. (2014), where the authors developed a large-scale of IoT infrastructure in a Santander city. Based on the city requirements, a user-friendly interface has been designed, and each citizen can test the IoT platform under different scenarios. These scenarios include security, ICT and non-ICT services, smart services, and unified control center. However, the data collected from different IoT sources is not tested for future urban design and planning. Also, the user demands are dynamically changed and the IoT-based smart environmental is not matured for supporting smart services such as smart buildings, and smart homes for the following two reasons: (1) compatibility of new technologies and optimization techniques (e.g., an IoT component such as wireless sensor network (WSN) has high packet loss in a heterogeneous wireless environment) and (2) limitation of current IoT-based solutions to specific application domains. In addition, different smart services such as smart transportation, smart waste management, and noise pollution are not reflecting a standard solution (Li et al. 2009; Maisonneuve et al. 2009; Nuortio et al. 2006).

IoT technology is not enough for the efficient design of a smart city. The elaboration of the data from different sources, which is obtained with the introduction of the Big Data technology, is required. Big Data supports off-line processing and can therefore help for designing and planning smart city and smart services. However, Big Data does not support real-time decisions. Various techniques based on Hadoop framework are implemented to analyze the data for better designing and planning smart city services. For example, a platform called City Data Analytics Platform (CiDAP) has been proposed (Cheng et al. 2015), where a layered platform of data processing between the data sources and applications has been developed. The entire platform consists of different components such as IoT agent, IoT collection unit (IoT broker), a communication city server, and a Big Data processing module. The data are collected from different sources and passed on to the IoT broker. The IoT broker separates the data based on different sensors IDs and then assigns an index number. Finally, the IoT broker sends data to IoT agent for further processing.

Similarly, several projects are developed based on Big Data analytics such as FIWARE (2018) and SCOPE (2018). The projects provide different mechanisms for the big data in real-time environment; though they are not open, they are available for use in different environments. From the literature, we can found example of Big Data analytics for the designing and planning smart city services (Li et al. 2009).

However, real-time elaboration and processing of large amount of data is still an interesting research challenge. In general, a smart city can be built based on real-time big data analytics and ICT/non-ICT communication models. The above literature identifies some important open issues that need to be addressed, for example, ICT/non-ICT communication models, real-time big data analytics, and acquisition and collection of data from different sensors installed into the smart city. In this chapter, we identify the need for an efficient model for smart cities based on Big Data architecture.

Proposed Scheme of the Big Data and Smart City

In this section, the development of a layer architecture that supports and integrates different data from IoT devices or City services is presented. City services produce a huge amount of data such that it is very difficult with traditional IT systems to collect and extract meaningful insights from. To address this problem, big data processing technologies have provided platforms for data distribution, processing, and interactive visualization (Kang et al. 2016). The proposed scheme is composed of different layers that enable the integration of smart city services with big data technologies. The proposed scheme consists of four levels: (1) infrastructure layer, (2) data orchestration layer, (3) services enablement layer, and (4) application layer. A brief overview of the proposed smart city architecture is provided in the next subsections followed by a description of the four levels of the proposed scheme (Fig. 9.1).

Infrastructure Layer

The infrastructure layer is the set of connected devices and the data generated and acquired from different sources. A smart city includes a complex and comprehensive applications domain to support the amount of data collected. The implementation of the smart city relies on all kinds of data

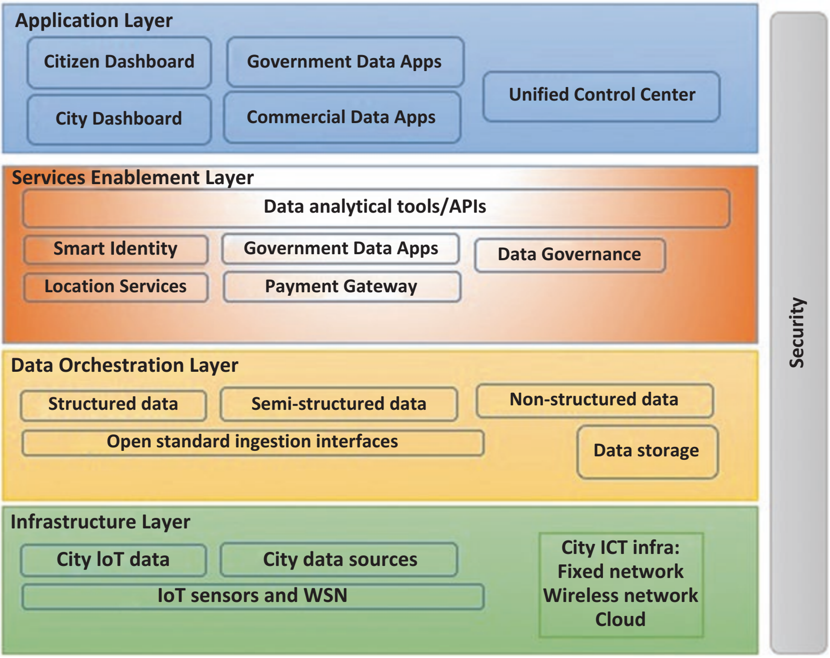

Fig. 9.1 Proposed scheme for Smart City, IoT, and Big Data (Source: Authors)

and computations due to their indispensability (Rong et al. 2014). The smart city aims to provide better life and services to the tenants, to optimize residential resources, to reduce traffic congestion, improve e-government and healthcare services, and to manage waste, among others. To achieve all the preceding aims, data acquisition on a daily operation basis is required. However, the data acquisition has become challenging due to the huge amount of data created by connected devices and people. For the further processing and the conversion into digital data, big data mechanisms are necessary to be implemented.

The installation of low-cost sensors is required for the acquisition of heterogeneous data from the city IoT. The city becomes smarter with the increased number of installed and connected sensors (Rong et al. 2014). In addition, there are other sources of data coming from government systems, commercial systems, as well as archived data that need to be revisited.

Another component of the infrastructure layer is the connectivity using multiple communication technologies, such as 5G, Wi-Fi, Bluetooth, and ZigBee, to transmit data from IoT devices to the data orchestration layer.

Data Orchestration Layer

The data orchestration layer acts as the mediation between the data acquisition, service, and application layers. Since all the processes such as data aggregation, manipulation, ingestion, transformation, and storing are performed in this layer, data orchestration is considered as the brain of the proposed architecture.

This layer collects all the unstructured data from the IoT and city sources, which are then stored in a shared data storage located either in a data center inside the smart city or in a cloud storage such as Amazon EWS, Openstack, S3, Google cloud services, or Microsoft Azure platforms. In order to perform the tasks, multiple hardware and software components are utilized in this layer. Stored data are processed using batch-based programming tool such as MapReduce (Dean and Ghemawat 2008) or other processing engines for big data platforms.

Additional platforms and technologies for data storing and manipulation are Hadoop distribution file system (HDFS), HBASE, and HIVE. In stream processing, data is processed and utilized quickly in order to properly support any changes in a smart city in real time. There are many technologies that can be used for real-time streaming processing of unstructured data such as Spark, Storm, and S4 (Neumeyer et al. 2010).

Services Enablement Layer

The service enablement layer provides the city services to the tenants in the smart city. The layer consists of multiple components such as smart identity, payment gateway, and smart home automation services. Smart identity services are introduced for better control, management, and visibility of activities. The prime focus of introducing smart identity is to maintain a single customer identity using various technology platforms, which all converge to the same subscriber of services. The following technologies will be considered: smart cards, magnetic cards, RFID tags, NFC enabled devices, and mobile applications in order to support smart identity. The smart identity becomes the key identity for all permanent or temporary subscriber/user of the smart services including access control, entry/exit to authorized buildings, virtual wallets, loyalty program management, and location-based services.

For these services, visitors received a temporary smart identity card while entering the complex and their information (name and national ID) are attached to it. There is an option online to load a balance for making some payments for city services from this smart card. While exit from the complex the card ID will be detached from the visitor information.

Another value-added service from this concept is to register all tenants and visitors in a loyalty program where users can redeem and consume points according to a present pointing scheme based on purchases from stores, restaurants inside the smart city. In addition, smart identity concept is the enabler to many location-based services that could be introduced such as push notifications to customers on their cell phones while roaming in the smart city campus.

With the home automation service residents are able to control or monitor signals from different appliances or basic services using IoT sensors. A smartphone or web browser can be used to control or monitor the home automation system. The home automation components provide security, air condition control, digital signage, smart lighting, and audio/video control.

Application Layer

The application layer is responsible for the data presentation. The main components are the unified control center, city dashboard, big data, open data, and government and commercial applications. The unified control center collects all the data from different smart services into the same platform. It integrates and interchanges between the other components all the required information. Such data can be used by the big data analytics model in order to predict behavior in the business and management performance contexts. The outcome of the analysis can be presented in a form of report. Dashboards are often used for easy interaction with the models. The ideal architecture requires a security model throughout the processes and examines security issues from a systems perspective to provide business value into the smart city organization.

Big Data Collection Process Overview

Smart city applications producing continuous large data from heterogeneous sources and the existing traditional warehouse approach is inadequate to handle such large amounts of data given the limited processing speed and the high storage cost. To address this problem, a smart hub approach is used for this chapter. The main concept of a data hub is to make different types of data readable after some processing and then available through multi-option interfaces.

Traditional Data Warehouse Approach

The traditional data warehouse approach collects all data into a centralized place as shown in Fig. 9.2. Afterward, extraction, transformation, and loading (ETL) tools allow processing of the data with the aim to identify relationships and build data model. Finally, data marts for each business unit can be identified and loaded. The main disadvantage of this approach is that it requires a considerable investment in the first place and even more to add new data sources or change the model. This approach is not efficient in real-time (read/write) operations on stored data.

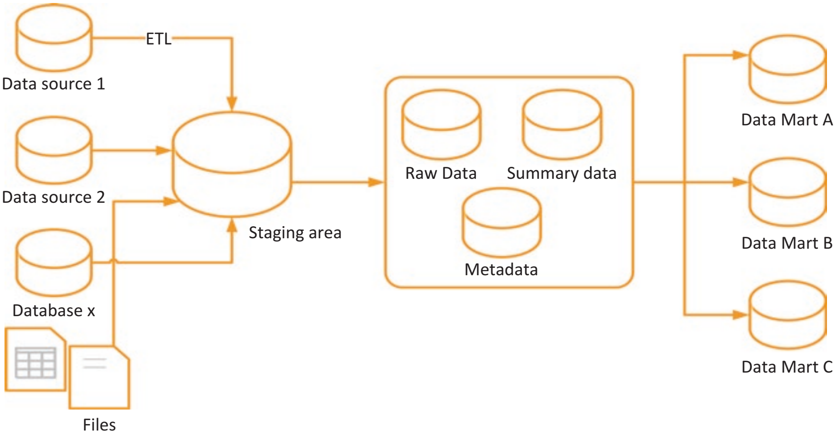

Fig. 9.2 Traditional data warehouse approach

Data Virtualization Approach

The main idea behind this approach is to have a data processing pipe that maps the data on the fly to data models represented into a controlled access data as shown in Fig. 9.3. This is similar to the processing virtualization concept. End users see what seems to be a data mart but that only carries a cached version of the data while the rest is pushed on the fly (data streaming).

The main advantage of this approach is that data maintaining functionality will not be replicated as each system maintains its own data while the user still can see specific selections of such data. This also reduces the level of investment and speeds up the time to market.

The main disadvantage here is that the performance of the system, as well as data availability, depends heavily on the level of integration of all the sources. This means that all systems should be ready all the time for data extraction. Based on the above discussion, a hybrid approach can be optimized. First, data virtualization is applied so as to monitor the types of data with a high request frequency for storing in the storage area. This storage is evolving with the size of data stored which gives the chance to minimize the initial investment and optimize the ongoing one to match the data demand.

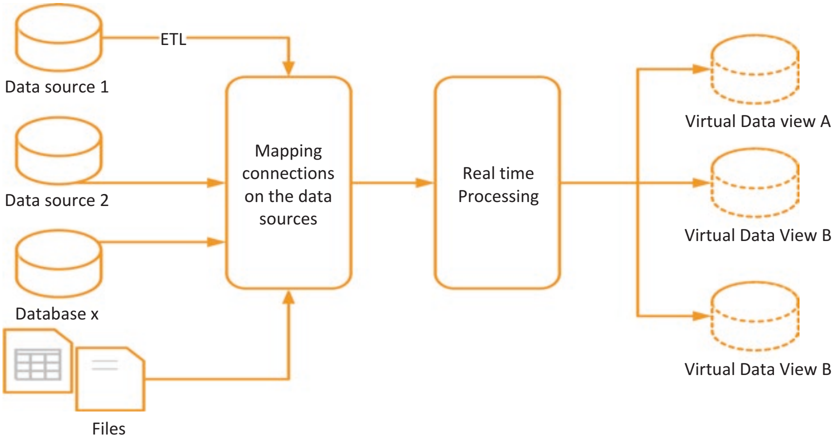

Fig. 9.3 Data virtualization approach

Applications of IoT and Big Data in the Smart City

Throughout the world, real estate is going through a major transformation. New cities are being created where the network and services running through them are integral to the design and day-to-day use. Technology will improve life for the office worker and resident alike, enhance building performance and conserve energy. The vision of the smart city is an environment designed to attract knowledge and talent from around the world and to spur innovation and creativity. The goal of the smart city is to promote education, art, science, medicine, industry, environment conservation, transportation, social communications, and public administration to enhance urban life.

In smart cities, people will move around, learn, consume, and interact through the use of advanced technologies that have a social dimension, are location-aware, mobile, and reactive. Smart cities are emerging partly because the oil wells that are funding the buzzing Middle East industry are not bottomless: governments in the Gulf are looking for new wealth generated from commerce and tourism to pay back their investment in the new smart built environment. It has become clear to developers and investors that technology can be a key differentiator to put clear blue water between them and their competitors. The truly smart city is not just a dream: the technology already exists. The challenge is to pull it all together into a unified, easily managed environment.

A great view or a prime site helps, but in a building boom that has brought unprecedented competition for landlords and property developers, there is one other essential for attracting new business: access to open data and broadband services, on every floor of every building, in homes and offices, shopping malls, and civic centers. Now real estate investors have a new option. In addition to selling units and facilities within the real estate campus, they can also deliver the converged broadband telecommunications services as the fourth utility (after electricity, gas, and water) and gain revenue from the services occupiers want and they can enjoy a higher return on their investment and achieve higher sale prices for their properties.

In a smart city, built-in broadband infrastructure is as important as the roads, supporting entertainment, information, and business communications. A smart city requires a high capacity backbone network to support future demand: and fiber terminated on each floor of every building so that whenever and wherever services are needed they can be simply and quickly activated. Using IP technology, a fiber network will carry information systems to facilitate the automation, centralization, and remote monitoring of essential building services, making these cities easier to live in, cheaper to operate and more energy efficient. Additional services can then be accessed without digging up roads or rewiring.

Fiber is just the foundation for a range of advanced services that can be layered on top. Customers of all types, including residents, facilities managers, retailers, corporates, small and medium businesses, and hoteliers can share the same single converged IP infrastructure supporting video, voice, and data services for different terminals over mobile and fixed line access technologies across an entire development. In the smart cities, there is a big ecosystem consisting of different types of buildings such as smart homes, smart buildings, and smart communities.

Smart Homes

In a smart city, residents are able to connect to the network instantly, accessing millions of games, films, music tracks, and countless other services. They can turn on the air conditioning from their mobile phone or check the family is safe by remotely accessing a home set-top camera. Assistive technology can give greater independence to the elderly or disabled. A home hub could wirelessly link multiple personal computers (PCs) and personal digital assistants (PDAs) to high-definition television, video, and radio, with multimedia services like online gaming and music downloading available on demand. The smart city extends all of the home automation capabilities to the outside of the home and your profile can follow you to any terminal within the smart city.

Smart Businesses

In a smart city, multisite businesses can network voice, video and business-critical enterprise applications. With the help of an international carrier, these can run seamlessly across corporate wide-area networks on different continents. Example applications are intranets and secure collaboration, email and voice mail, conference calls and instant messaging. Unified communications make it easy for employees to hot desk or work remotely. Managed firewalls, antivirus, storage area networking, remote access, messaging, and web hosting can all be added. Branches of multinational businesses will have all of the powerful communication tools that offices in London, New York, or Tokyo have.

Smart Hotels

Hotels in a smart city can offer guests video-on-demand and IPTV as well as office services like video-conferencing, which can be billed on a per-use or monthly basis. In the public spaces, plasma/LCD televisions, wireless access, digital signage, and virtual reception can be provided. Add to these, point-of-sale terminals in bars and restaurants, a front office booking suite, automated check-in and smart-card room locks. Room comfort settings can be delivered over IP phones and smart bathrooms let management know when supplies run short.

Smart Shopping Malls

Malls and souks can use multiple VoIP phones and business applications on corporate intranets to check stock availability. Applications can stream video feeds onto in-store display screens and security cameras as well as connect to chip-and-pin machines and electronic payment terminals.

Smart Buildings

Office buildings already contain a number of networks: fire alarms, security alarms, door access controls, utilities monitoring, lighting systems, lifts, heating, ventilation, and air conditioning. This is a complex environment with high installation costs and limited automation. The aforementioned services can all be run over the same network, delivering better performance for less cost. Accurate information gives owners the ability to save energy by offsetting peak loads. Intelligent buildings know when to turn off lights or turn down the cooling system and even before they are completed, developers are able to adjust internal temperatures for the comfort of building workers.

Smart Government and Public Services

The need for physical human resources and dedicated resources for management of offices can all be reduced by providing online access to voting, birth certificate applications, tax payments, and so on. On-campus hospitals can be created with telemedicine in mind—outpatients can have their heart rates, blood pressure, and insulin levels monitored and recorded remotely with the vital statistics fed into telemedicine control systems.

Schools and universities in a smart city can support distance learning and rich media. Lectures delivered by professors around the world can be conducted via telepresence. In a smart city, smart government and public services mean the public are healthier, better educated, and disasters can be more easily prevented or mitigated. Even the crime rate can be ultimately reduced.

Smart Communities

Smart communities are cities and corporate facilities that leverage technologies to better engage with citizens. Smart technology helps public- and private-sector planners develop viable platforms made up of IoT and M2M technologies. Smart technologies help unify systems and use resources more efficiently and address community challenges like high-energy costs, traffic congestion, and public safety.

With smart communities and cities, one can collect critical data and gain near real-time visibility into vital operations. This helps in streamlining the decision-making process and facilitating faster response. For example, if a water pipe breaks, sensors will be able to detect the sudden pressure change and notify public works officials, who could then dispatch the closest field crew to repair the damage. In the event of area flooding, residents could receive alerts, traffic lights could be adjusted to steer drivers toward safety and digital signage could be used for information sharing. This is one of the many ways in which smart cities' technology can help communities become more resilient.

When a city is able to collect data from IoT and connected machine technologies, communities can benefit from revenue generation opportunities, make the use of efficient limited resources, and attract businesses, residents, and workers.

Big Data Tools and Systems

Big data tools and systems are required to implement the functionality and services for the smart city. The approached list of vendors that allow us to achieve the required design with their various constraints are presented below.

Hortonworks Framework

The Hortonworks framework is the main platform for the Hadoop infrastructure and Big Data analytical stack as shown in Fig. 9.4. The Hortonworks framework is divided into three main layers:

(a) The ingestion layer: We will use Sqoop to ingest data from structured query language (SQL) and structured data sources. Then, we use Apache Flume to ingest real-time and change-based triggering data. Finally, we use Kafka to ingest big non-structured data, volumes up to terabytes.
(b) The data management layer: We use Apache Hadoop file system with YARN for data operation management and jobs handing. YARN is the architectual center of Hadoop that allows multiple data processing engines such as real-time streaming, interactive SQL and batch processing to handle data stored in a single platform. We utilize HDFS V2.2 which is a scalable, fault-tolerant, distributed storage system that works closely with a wide variety of concurrent data access applications and is coordinated by

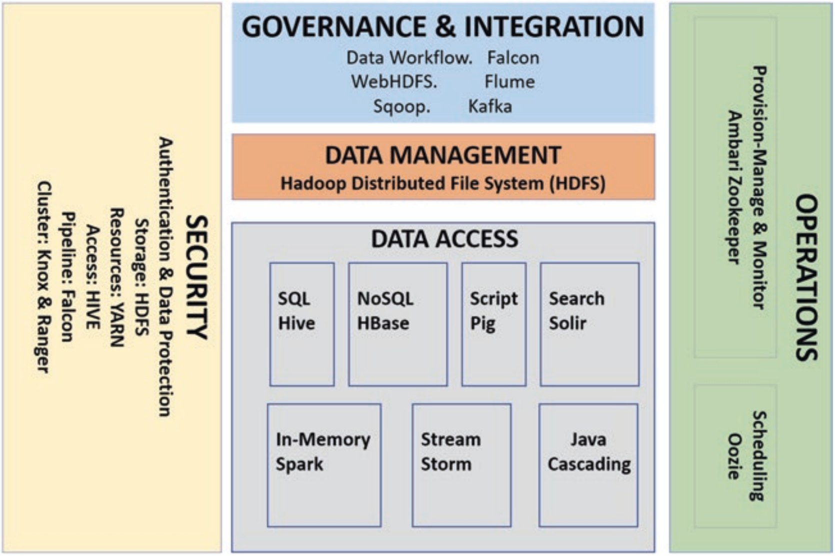

Fig. 9.4 Hortonworks proposed framework

YARN. HDFS will "just work" under a variety of physical and systemic circumstances. By distributing storage and computation across many servers, the combined storage resource can grow linearly with demand while remaining economical at every amount of storage.

(c) The analytical and data access: In this layer, tools that allow multiple techniques for data access are used, like: Data flow Scripting language (Pig Latin) & Java scripts data access using Cascading. Interactive SQL access (will use Hive for large & distributed SQL-based data access while using HAWQ "which became open source as well now" for low latency interactive SQL like data access like interactive dashboards in our case). HBase for value store high volume noSQL data access and Accumulo for secured noSQL data access. Storm & Spark: Storm for Data streaming & Spark for in memory data access.

The Hortonworks frame utilizes many other tools for data management that we intend to use, such as:

Falcon for data life cycle management as Falcon simplifies the development and management of data processing pipelines with a higher layer of abstraction, taking the complex coding out of data processing applications by providing out-of-the-box data management services. This simplifies the configuration and orchestration of data motion, disaster recovery, and data retention workflows.

Apache Ranger for data access security on HDFS level as it offers a centralized security framework to manage fine-grained access control over Hadoop data access components like Apache Hive and Apache HBase. Using the Apache Ranger console, security administrators can easily manage policies for access to files, folders, databases, tables, or column. These policies can be set for individual users or groups and then enforced within Hadoop.

Knox as the data access application programming interfaces (APIs) gateway as it provides perimeter security so that the enterprise can confidently extend Hadoop access to more of those new users while also maintaining compliance with enterprise security policies.

Ambari for the cluster management as Ambari enables system administrators to provision, manage, and monitor a Hadoop cluster, and also to integrate Hadoop with the existing enterprise infrastructure. Some of its main features are provisioning, managing, monitoring a Hadoop cluster, and integrating Hadoop with the Enterprise.

Apache ZooKeeper that provides operational services for a Hadoop cluster. ZooKeeper provides a distributed configuration service, a synchronization service, and a naming registry for distributed systems. Distributed applications use Zookeeper to store and mediate updates to important configuration information.

Apache Oozie as it is a Java Web application used to schedule Apache Hadoop jobs. Oozie combines multiple jobs sequentially into one logical unit of work. It is integrated with the Hadoop stack, with YARN as its architectural center, and supports Hadoop jobs for Apache MapReduce, Apache Pig, Apache Hive, and Apache Sqoop. Oozie can also schedule jobs specific to a system, like Java programs or shell scripts.

IoT Intelligent Application Enabler Platform (IAE)

The Intelligent Apps enabler is a flexible platform that facilitates and rationalizes the exchange of data within any IoT solution, fulfilling the customer needs for device, communication, and message management. It eliminates the usual complexity of developing and integrating IoT applications. It is a generic platform allowing us to create and manage end-to-end IoT services. With this platform we can deliver a purpose-built end-to-end service beyond connectivity, encompassing terminal device, and network and back-end applications.

One of the main goals of this platform is to enable as much as possible the utmost replication from one project to another and then optimize the costs. Focus is on building reusable blocks to provide end-to-end solutions to multiple customers. This platform will fulfill the IoT needs of device management, communication and message management, business data management, and assets repository management.

IoT Gateways

The IoT gateway offers a key building block to enable the connectivity of legacy industrial devices and next-generation intelligent infrastructure to the IoT. It integrates technologies and protocols for networking, embedded control, enterprise-grade security, and easy manageability on which application specific software can run. IoT Gateways enable:

- Connectivity up to the cloud and enterprises.
- Connectivity down to sensors and existing controllers embedded in the system.
- Pre-process filtering of selected data for delivery.
- Local decision-making, enabling easy connectivity to legacy systems.
- A hardware root of trust, data encryption, and software lockdown for security.
- Local computing for in-device analytics.

Cloud Infrastructure

The enterprise-class UCS C240 server extends the capabilities of the UCS portfolio in a 2RU form factor. Based on the Intel® Xeon® processor E5-2600 v3 series, it delivers good combination of performance, flexibility, and efficiency. In addition, it comes with good levels of internal memory and storage expandability with exceptional performance.

OpenStack Cloud Platform

OpenStack is the leading open cloud platform. Ubuntu is the reference operating system for the OpenStack project, which is why deploying OpenStack with Ubuntu is the best way to ensure a straightforward implementation. Ubuntu Openstack contains all the current integrated OpenStack projects and some additional technologies beneficial to helping run an OpenStack cloud. The following is a high-level overview of our reference implementation for production use:

Compute—Nova: Nova, an OpenStack component enables the provision and management of large networks of virtual machines, creating a redundant and scalable cloud-computing platform. It gives the needed to run instances, manage networks, and control access through users and projects. Like the rest of the Ubuntu operating system, it supports most standard hardware configurations and well-known hypervisors.

Storage—Swift or Ceph: Swift, OpenStack Object Storage creates redundant, scalable object storage using clusters of standardized storage servers. Rather than a file system or real-time data storage system, it provides a long-term storage system for more permanent, static data. Examples include virtual machine images, photo storage, email storage, and backup archiving. Ceph is used for block storage as part of OpenStack Cinder. Ceph provides a feature-rich experience and is also able to provide Object Storage via a gateway giving an option to standardize on a single storage technology based on storage requirements.

Image service—Glance: The OpenStack Image Service provides discovery, registration, and delivery services for virtual disk images. It includes a standard REST interface for identifying them in back-end stores such as OpenStack Object Storage, with new virtual disk images being registered via the Image Service. Administrators can also access information on publicly available disk images and use the client library for streaming virtual disk images.

Authentication—Keystone: Keystone, the OpenStack Authentication service provides identity, token, catalogue, and policy services for use by OpenStack components. It provides a pluggable back-end that has been designed to support various protocols (e.g. Basic Auth, OAuth, OpenID, PKI) for authentication and authorization, allowing clients to obtain security tokens to access different cloud services.

Management—Horizon: The OpenStack management service or dashboard provides OpenStack users with a web-based user interface with which to control OpenStack's component services (Nova, Swift, Keystone, and Glance) and a single API with which to access them.

Networking—Neutron: Neutron, the networking service provides an API that can define network connectivity and addressing in the OpenStack cloud. Neutron enables operators to leverage different networking technologies supported by Ubuntu OpenStack such as Juniper's Contrail, Nuage Networks, MidoNet, VMware NSX, PlumGrid, CPlane Networks, and Calico from Metaswitch. Standalone, Neutron provides an API to configure and manage a variety of network services ranging from L3 forwarding and NAT to load balancing, edge firewalls, and IPsec VPN.

Metrics—Telemetry: Previously known as Ceilometer, the Telemetry module collects measurements about the OpenStack system and stores it in the form of samples in order to provide data about anything that can be billed. In addition to system measurements, Telemetry also captures event notifications triggered when various actions are executed in the OpenStack system. This data is captured as Events and stored alongside metering data.

Juju and MAAS: Juju and MAAS are the fastest and most integrated way to deploy OpenStack on Ubuntu. Using libraries of "charms" developed from experiences at customer sites makes it simple to deploy, configure, and scale out cloud services with only a few simple commands. MAAS is the bare-metal provisioning tool that turns hardware environment into a cloud in minutes. It takes the pain out of detection and configuration and gets the servers ready for Juju.

Virtualization: Ubuntu Server includes open-source hypervisors LXD/LXC alongside its default option, KVM. All are supported as virtualization options for Ubuntu OpenStack deployments.

jBilling and PayOne Payment Gateways

jBilling is a web-based enterprise billing and rating system. It manages subscribers with automatic invoicing (email and PDF) and payment processing (credit cards, checks, and direct deposit). It is robust, well documented, and easy to use, while PayOne is an unrivaled transaction processing solution that authorizes multiple payment methods (credit and debit cards) through different payment channels such as e-commerce, mobile, MOTO, and IVR. With a wide range of features, it provides both web and mobile experience to cardholders to conduct their payment transactions. Through advanced Fraud Detection and Prevention, PayOne gateway helps in minimizing the hidden costs associated with fraud and chargebacks.

QlikView Interactive BI

QlikView Interactive BI is a business discovery platform that delivers self-service BI that empowers business users by driving innovative decision-making with many features such as:

- Consolidating relevant data from multiple sources into a single application.
- Exploring the associations in the data.
- Enabling social decision-making through secure, real-time collaboration.
- Visualizing data with engaging, state-of-the-art graphics.

- Searching across all data—directly and indirectly.
- Interacting with dynamic apps, dashboards, and analytics.
- Accessing, analyzing, and capturing data from mobile devices.

Identity and Access Management

Manage the end-to-end lifecycle of user identities across all enterprise resources, both within and beyond the firewall and into the cloud. The Identity Management platform delivers scalable solutions for identity governance, access management, and directory services. This modern platform helps organizations strengthen security, simplify compliance, and capture business opportunities around mobile and social access.

API Gateway

API gateway allows integrating cloud services, on-premise services proxy and manages interactions with cloud services. In addition, it can secure and manage APIs and SSO for Web Services Extend Enterprise Security to mobile and cloud applications. API gateway extends authentication, authorization, and risk policies to mobile, cloud, and enterprise applications. It enforces Data Security Policies through scanning messages and data payloads against security and privacy policies and block, redact, remove, or encrypt data.

Research Challenges

The big data context in the smart city has opened a number of opportunities for a new value proposition (Lohr 2012). In this section, some business and technology challenges are presented.

Business Challenges

Many smart city business models have been presented from commercial companies for a business market growth using new technologies such as IoT and Big Data. During the master planning development, the real estate developers faced difficulties for the application of Big Data into to the existing planning. Another challenge is the sustainability of cities using IoT and Big data technologies to improve their services. It is also in general challenging to clearly recognize the benefits of the IoT and Big data usage. The integration cost for the smart city is another challenge worth mentioning. Different hardware and software platforms are typically required for integration in order to effectively provide smart services using Big Data tools and systems. A key to reducing the cost is the usage of robust open standard technologies and frameworks.

Technology Challenges

The rapid growth of the smart services in the smart city and the huge amount of data requires a well-defined controlling and monitoring technology mechanisms. Privacy is an important challenge since private information of citizens may be exposed to analysis, sharing and use by third parties. Moreover, smart cities technologies raise a number of cybersecurity concerns that require attention especially when IoT technology is used.

Effective data analytics and data integration is another technology challenge worth mentioning. Despite advancements in integration efforts and applications, data quality remains a barrier to any data integration mechanism. This is particularly the case when data are incorrect, missing, use the wrong format and/or are incomplete (Gouveia et al. 2016).

Artificial intelligence and machine learning algorithms can be efficient in addressing many of the data-related challenges identified above, (Jin et al. 2014). However, their ease of applicability for extremely large data sets, such as in the case of smart cities, needs to be further evaluated (Tsai et al. 2014).

Conclusion

In this chapter, a comprehensive view of the role of IoT and Big data in a smart city is presented. In this context, different technologies and open-source software required for building the Big Data stack are discussed. Moreover, a scheme for the integration of Big Data, IoT and smart services under the smart city area is proposed. Finally, several business and technical challenges are identified for future research.

References

Ashton. (2009). That Internet of things. *RFID Journal* [online], http://www.rfidjournal.com/article/print/4986, last accessed 15/4/2017.

Cheng, B., Longo, S., Cirillo, F., Bauer, M., and Kovacs, E. (2015). Building a big data platform for smart cities: experience and lessons from santander, *in*

Proceedings of the 4th IEEE International Congress on Big Data (BigData Congress '15), pp. 592–599, New York, NY, USA.

Dean, J., and Ghemawat, S. (2008). MapReduce: simplified data processing on large clusters, *in Communications of the ACM*, 51(1), 107–113.

FIWARE. (2018). Open Source Platform. *FIWARE*, https://www.fiware.org/, last accessed 15/5/2018.

Gouveia, P., Seixas, J., and Giannakidis, G. (2016). Smart City Energy Planning: Integrating Data and Tools, in *25th International Conference Companion on World Wide Web*, Montreal, QC, Canada.

Jin, J., Gubbi, J., Marusic, S., and Palanoswami, M. (2014). An information framework for creating a smart city through internet of things, *in Internet of Things IEEE*, 1 (2), 112–121.

Kang, Y., Park, I., Rhee, J., and Lee, Y. (2016). MongoDB-Based Repository Design for IoT Generated RFID/Sensor Big Data, *in IEEE Sensors Journal*, 16 (2), 485–497.

Li, X., Shu, W., Li, M., Huang, H. Y., Luo, P. E., and Wu, M. Y. (2009). Performance evaluation of vehicle-based mobile sensor networks for traffic monitoring, *IEEE Transactions on Vehicular Technology*, 58 (4), 1647–1653.

Lohr, S. (2012). The age of Big Data, *New York Times*, Feb. 11, 2012.

Maisonneuve, N., Stevens, M., Niessen, M. E., Hanappe, P., and Steels, L. (2009). Citizen noise pollution monitoring, in *Proceedings of the 10th Annual International Conference on Digital Government Research: Social Networks: Making Connections between Citizens, Data and Government*, Puebla, Mexico.

Neumeyer, L., Robbins, B., Nair, A., and Kesari, A. (2010). S4: Distributed Stream Computing Platform, *in IEEE International Conference on Data Mining Workshop (ICDMW)*, New York, NY.

Nuortio, T., Kytojoki, J., Niska, H., and Braysy, O. (2006). Improved route planning and scheduling of waste collection and transport, *Expert Systems with Applications*, 30 (2), 223–232.

Rong, W., Xiong, Z., Cooper, D., Li, C., and Sheng, H. (2014). Smart city architecture: a technology guide for implementation and design challenges, *China Communications*, 11 (3), 56–69.

Sanchez, L., Munoz, L., Galache, J. et al. (2014). SmartSantander: IoT experimentation over a smart city testbed, *Computer Networks*, 61, 217–238.

SCOPE. (2018). A Smart-city Cloud-based Open Platform and Ecosystem, *SCOPE*, https://www.bu.edu/hic/research/highlighted-sponsoredprojects/scope/, last accessed 15/5/2018.

Tsai, W., Lai, C., and Vasilakos, A. (2014). Future Internet of Things: open issues and challenges, *in Wireless Networks*, 20 (8), 2201–2217.