

# Big Data, Internet of Things and Cloud Convergence for E-Health Applications

George Suciu, Victor Suciu, Simona Halunga, and Octavian Fratu

Faculty of Electronics, Telecommunications and IT, Telecommunication Department,  
University POLITEHNICA of Bucharest, 060071, Bucharest-6, Romania  
{george.suciu,victor.suciu}@radio.pub.ro,  
{shalunga,ofratu}@elcom.pub.ro

**Abstract.** Big data storage and processing are considered as one of the main applications for cloud computing systems. Furthermore, the development of the Internet of Things (IoT) paradigm has advanced the research on Machine to Machine (M2M) communications and enabled novel tele-monitoring architectures for E-Health applications. However, there is a need for converging current decentralized cloud systems, general software for processing big data and IoT systems. The purpose of this paper is to analyze existing components and methods of integrating big data processing with cloud M2M systems based on Remote Telemetry Units (RTUs) and to propose a converged E-Health architecture built on Exalead CloudView, a search based application. Finally, we discuss the main findings of the proposed implementation and future directions.

**Keywords:** Big Data, IoT, M2M, Cloud computing, E-Health.

## 1 Introduction

Cloud computing, Big data and Internet of Things (IoT) are popular ICT paradigms and their features can be combined for shaping the next generation of eHealth systems.

Big data in E-Health enables the transformation from hypothesis driven research to data-driven research by processing large volumes of heterogeneous medical data [1]. Furthermore, by using a search based application it is possible to find weak signals in big data [2]. In particular, given the E-Health application, it is possible to leverage trivial and non-trivial connections between different sensor signals and existing big data, in order to find new ways to provide remote diagnostics, better understanding of diseases and development of innovative solutions for therapeutics.

The aggregation and analysis of such weak signals will provide evidence of connections between health problems and environmental related issues faster and better than trivial mining of sensor data. As a consequence, the software has a significant potential for matching E-Health applications and critical challenges that are related in non-obvious ways in IoT scenarios.

Subsequently, the IoT paradigm relies on the identification and use of a large number of heterogeneous physical and virtual objects (i.e. both physical and virtual representations), which are connected to the internet [3]. IoT enables the communication between different objects, as well as the in-context invocation of their capabilities (services) towards added-value applications. Early IoT applications are based on RFID (Radio Frequency Identification) and Wireless Sensor Network (WSN) technologies and deliver tangible benefits in several areas including manufacturing, logistics, trade, retail, green/sustainable applications, as well as other sectors.

In the case of M2M, important research initiatives are proposing new reference models, defining standards and new communication architectures, with different approaches for solving security, reliability and energy-efficiency problems. However, many efforts are focused on largely distributed critical infrastructures and just a few initiatives are dedicated to the definition of platforms for deployment and execution of new M2M applications, based on new generations of WSN and innovative sensor mining methods that can be deployed according to a cloud computing model [4].

Indeed, M2M applications are envisioned to be hosted within cloud computing systems and contribute to the convergence of interactions between the real and virtual world, thus accomplishing improvements in industrial productivity and quality of life for citizens, based on secure and dependable automation of sensing and actuating tasks.

However, we argue that this kind of approach is sub-optimal for M2M because of its inherent paradigms: M2M applications are highly autonomous, implement simple and repetitive interactions in highly constrained environments, with limited scope in time and space, and must respond in a reliable manner to mission critical QoS needs, whereas usual cloud applications require that big volumes of data are redundantly stored and provided through content delivery networks, to be available for very long periods and accessible as ubiquitously as possible for the use by human beings. Also, we discuss the applicability for other information sources and future use case scenarios.

At the same time, the cloud computing paradigm [5] realizes and promotes the delivery of hardware and software resources over the Internet and according to an on-demand utility based model. These services hold to promise to deliver increased reliability, security, high availability and improved QoS at an overall lower total cost of ownership (TCO).

The paper is organized as follows. Section 2 presents an overview of the state of the art in Cloud computing, IoT, Big data and M2M. Section 3 analyzes the components of a convergent IoT - Cloud computing architecture for E-Health applications, while Section 4 presents Big data processing tools integrated in the previously mentioned architecture. Finally, Section 5 draws the conclusions and presents future directions.

## 2 Related Work

Starting with the first versions of Cloud computing, IoT and Big Data technologies, it has become apparent that their convergence could lead to a range of new applications, including in eHealth. Most IoT applications are based on M2M communication protocols between large numbers of heterogeneous geographically distributed sensors. As a result, they need to handle many hundreds (sometimes thousands) of sensor streams, and could directly benefit from the immense distributed storage capacities of cloud computing infrastructures. Furthermore, cloud infrastructures could boost the computational capacities of IoT applications, given that several multi-sensor applications need to perform complex processing that is subject to timing (and other QoS constraints). Also, several IoT services (e.g., large scale sensing experiments, smart city applications) could benefit from a utility-based delivery paradigm, which emphasizes the on-demand establishment and delivery of IoT applications over a cloud-based infrastructure.

The proclaimed benefits of the IoT/cloud convergence have (early on) given rise to research efforts that attempted to integrate multi-sensory services into cloud computing infrastructures. The most prominent of these efforts are the so-called sensor-clouds, which blend sensors into the data centre of the cloud and accordingly provide service oriented access to sensor data and resources [6]. Several recent research initiatives are focusing on real-life implementation of sensor clouds, including open source implementations [7].

In addition to research efforts towards open source sensor-clouds, there are also a large number of commercial on-line cloud-like infrastructures, which enable end-users to attach their devices on the cloud, while also enabling the development of applications that use those devices and the relevant sensor streams. These systems provide tools for application development, but they offer very poor semantics and no readily available capabilities for utility based delivery. There is also a number of other projects which have been using cloud infrastructures as a medium for Machine-to-Machine (M2M) interactions (e.g., [8]), without however adapting the cloud infrastructure to the needs of the IoT delivery.

In the area of IoT applications (e.g., for smart cities), we are also witnessing cases of IoT/cloud convergence, for example deploying sensor streams over cloud infrastructure, as means to benefitting from the cloud's storage capacity and applications hosting capabilities. A similar approach is followed in the scope of the FP7 SmartSantander project [9]. We note that smart cities is a privileged domain for exploring and realizing the IoT/cloud convergence, given that such applications need to manage and exploit a large number of distributed heterogeneous sensor streams and actuating devices.

The plethora of data types and formats (structured, unstructured, semistructured or multistructured) as well as the diversity of generating sources (sensors, devices, social networks, web content, mobile phones, etc.) generate a large variety of data. According to [10], multistructured data already represents 80% of the volume of data that is available in an organization.

The economic potential of Big Data represents the greatest challenge and it consists of finding value in the large volume of unstructured data in (near) real time. The tendency to use this information for business analytics is becoming a worldwide management practice.

Fruitful sources of data, recently come to attention (IoT, sensor networks, social networks) have given an unprecedented growth rate of the volume of available microeconomic and macroeconomic data. It is estimated that in the year 2012 alone, 2.5 ExaBytes of data have been generated daily [10].

M2M communications (or Machine Type Communications - MTC) can use either wireless or fixed networks, gaining in recent years a considerable interest among mobile network operators. M2M communication solutions that use mobile networks can prove to be easier to deploy and can support a large number of de-vices, most importantly those with mobility features. The ubiquitous connectivity nature of mobile networks (especially from 3GPP Rel. 8 onwards [11]) will enable M2M services that require reliable and immediate data delivery to distant M2M servers.

The different behaviour of M2M devices, compared to plain mobile network terminals poses a need for optimizing networking solutions, in order to specifically tailor them for M2M communications in mobile networks. Therefore, 3GPP, the Open Mobile Alliance (OMA), IEEE, ETSI and a number of other standardization bodies have launched standardization activities on M2M communications [11].

Regarding 3GPP, the focus is on system optimizations that prevent M2M signalling congestion and network overload. These are the main important issues that prevent the mass-market adoption of M2M services.

### **3 Components of the Converged Cloud – IoT E-Health Architecture**

To illustrate the need for interconnecting extremely heterogeneous sensors and gateways we consider the following example scenarios, which show short-term realistic applications of smart things enabled by M2M communication:

- Continuous care: Citizens affected by chronic diseases can be provided with sensors continuously monitoring relevant health parameters, which produce data that are conveyed through a smart phone to a remote centre for real-time analysis. In case a potentially dangerous situation is detected an alarm can be fed back to the smart phone.
- Ambient assisted living: Activity detection sensors installed in houses where senior or disabled citizens live send data to a remote centre, so as to generate alarms in case anomalies are detected with respect to a typical pattern (e.g., prolonged lack of movement during daytime).
- other scenarios: Smart grid, Traffic management, Fleet management

### 3.1 Cloud Architecture

In this section we will introduce SlapOS [12], providing an overview of the decentralized cloud system and are going in particular to explain the role of Master node and Slave nodes, as well as the software components which they rely on to operate a distributed cloud for E-Health applications.

Slave nodes request to Master nodes which software they should install, which software they show run and report to Master node how much resources each running software has been using for a certain period of time.

Master nodes keep track of available slave node capacity and available software. Master node also act as Web portals and Web service so that end users and software bots can request software instances which are instantiated and run on Slave nodes.

Master nodes are stateful. Slave nodes are stateless. More precisely, all information required to rebuild a Slave node is stored in the Master node. This may include the URL of a backup service which keeps an online copy of data so that in case of failure of a Slave node, a replacement Slave node can be rebuilt with the same data.

It is thus very important to make sure that the state data present in Master node is well protected. This could be implemented by hosting Master node on a trusted IaaS infrastructure with redundant resource. Or - better - by hosting multiple Master nodes on many Slave nodes located in different regions of the world thanks to appropriate data redundancy heuristic. We are touching here the first reflexive nature of SlapOS. A SlapOS master is normally a running instance of SlapOS Master software instantiated on a collection of Slave nodes which, together, form a trusted hosting infrastructure. In other terms, SlapOS is self-hosted.

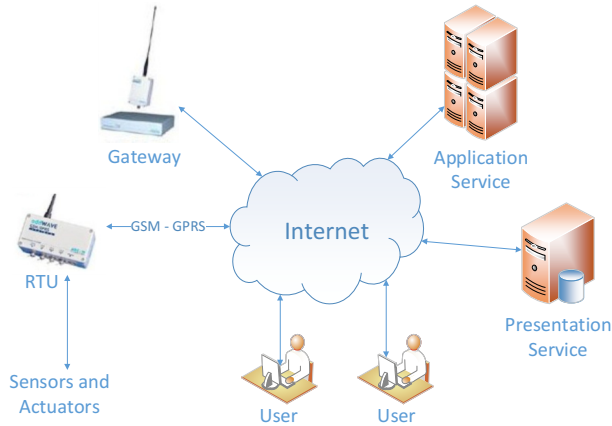
SlapOS Slave nodes are relatively simple compared to the Master node. Every slave node needs to run software requested by the Master node. It is thus on the Slave nodes that software is installed. To save disk space, Slave nodes only install the software which they really need.

### 3.2 M2M Architecture for E-Health

In Fig. 1 we present the general structure of the system that we propose for the tele-monitoring of health related parameters. At each of the monitored installation sites a system is set up composed mainly from distant RTUs, sensors and actuators. RTUs capable to communicate with the Gateway through GSM-GPRS and Internet will be used in standard configurations.

The system's key elements are:

- Gateway, which ensures the communication with the RTUs and available resource management;
- Presentation Service (PS) which is hosted on a computer with server features (for example, unattended operation 24/7), equipped with a software package focused mainly on data presentation in various forms, entirely available to users.
- Application Service (AS), focused on special tasks, which PS cannot perform.



**Fig. 1.** General structure of the E-Health system

Practically, all system communication is done through the Internet and this gives the system investment and mostly operational advantages. It is mentioned that the users can access the processed data, offered by the PS and AS anywhere and anytime, from any terminal with Internet access (PC, tablet, mobile phone etc).

The system's central elements are configured and scaled so that they would allow a system capacity of 100 RTUs.

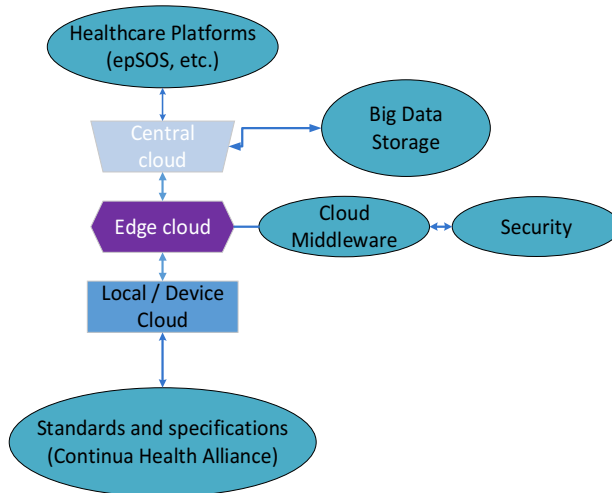
#### **4 Big Data Processing for Convergent Cloud IoT E-Health Applications – Characteristics and Discussions**

Currently there are many solutions on the market for search and analysis of large volumes of information. These solutions are focused on semantic technologies for aggregating and collating data, both structured and unstructured [13].

The cloud middleware layer can be developed based on three sub-layers, as follows: local cloud components on slave sensors node, edge cloud components on master gateway nodes and central cloud components on application / datastore server nodes. The resulting cloud architecture for the interconnection with a healthcare platform is illustrated in Fig. 2.

Cloud middleware enables heterogeneous devices to act as data sources and to integrate data from other healthcare platforms. Being located at the intersection between central and local cloud, the cloud middleware is designed to implement home gateway functionality and to ensure the security of user information. The local cloud enables the interconnection with different devices through standardized network protocols and interfaces.

In our approach we use CloudView Exalead [14], a search platform which offers wide access to information on infrastructure level and is used for both SBA (Search Based Application) for online and enterprise level. This application combines semantic technologies for developing applications such as drag-and-drop, as well as qualitative and quantitative analysis to provide information to the user.



**Fig. 2.** Big data processing on convergent IoT / Cloud E-Health architecture

Located at the intersection between search and Business Intelligence, CloudView is also a platform for the Exalead search engine, which was designed to implement semantic processing and selective navigation data volumes that are on the Web, making it easy for users searching and analysing information and enabling organizations to improve their knowledge and resources exploitation.

Exalead CloudView is a tool that allows the exploitation of large volumes of structured and unstructured data from different sources available to search applications, and enables their presentation in an intuitive search interface. CloudView uses semantic technology platform combining formats, structures and terminologies in a modular index that provides continuous access and use of resources through a server distribution and data redundancy.

CloudView accesses documents from different information sources using so-called connectors. Each connector uses protocol data source to connect their own information source and access the documents to be indexed. With connectors user can add new documents, update and delete existing documents, extract current list of indexed documents and manage data security.

CloudView provides multiple interfaces for data management and application configuration itself. These interfaces are represented by several APIs (Application Programming Interface), including a Push API for creating custom connector, a Search API with a role in the development of other applications Search and Management API for configuring and managing the processes of indexing and search. Push API allows data from any source to be indexed via CloudView and supports basic operations required for the development of new connectors, both managed and unmanaged. A connector is part of the managed code that runs in CloudView code that can be developed in Java. A connector is an external unmanaged sending data CloudView using Push API. An unmanaged connector can be developed in any

language or through customers' Push APIs (available in Java, C# and PHP) or through HTTP API.

Fig. 3 provides a simplified view of the indexing process. It works as follows: connectors accessing data sources convert files in documents and send them to the Push Server that divides documents into groups (analysis).

Analysis is performed sequentially, each document being processed. It makes retrieval of text, semantic processing, and other custom operations and location. The analysis also calculates the data to update the index. Once indexing is done and updated, the new documents are available for searching.

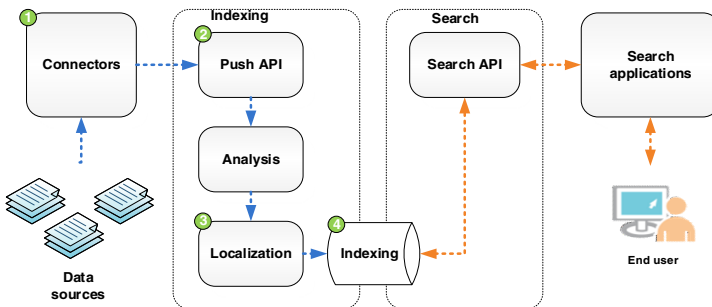


Fig. 3. Simplified view of the indexing process

Connectors operate early in the process of indexing of CloudView. They send documents from different sources to a Push API server through PAPI protocol. When creating new connectors the connector server that will host these connectors must be specified. If the connector server is already existing, all new connectors will be automatically associated with it.

With the advent of cloud computing, the convergence of the cloud computing with WSN has been attempted, as an extension of the sensor grid concept in the scope of on-demand elastic cloud based environments. In particular, the convergence of cloud computing with WSN aims at compromising the radically differences and conflicting properties of the two technologies. In particular, sensor networks are location specific, resource constrained, expensive (in terms of development/deployment cost) and generally inflexible in terms of resource access and availability. On the contrary, cloud based infrastructures are location independent provide a wealth of inexpensive resources, as well as rapid elasticity

Fine-grained (raw) data have to be conveyed in a centralized manner over the Internet from sensors up to the remote centre, so as to give the latter the high-resolution information it needs to take decisions. Therefore, the things and gate-ways are effectively separated from the back-end, which has storage and computation functions, both physically, through the Internet, and logically via a set of abstraction layers. Such an approach has worked well in the past for similar technologies, like the Web, from which the initial IoT attempts inherit most of their design and characteristics. However, we argue that full and sustainable exploitation of M2M applications needs a tighter integration between the real and virtual world. The



following issues of what we can call “the current approach” can be identified: scalability, security, reliability, QoS, resource/energy efficiency and multi-domain implementation.

To overcome the limitations of the current systems for M2M applications, we propose a novel approach based on the following principles:

1. Storage and processing of data are as close as possible, in space and time, to where they are generated and consumed.
2. Important non-functional requirements, namely security, reliability and QoS, are supported by means of a tight integration of the high-level services provided with the physical resources of the peripheral devices, i.e., things and gateways.
3. Energy efficiency and scalability of the systems are achieved through the distribution of on-the-spot inferred content, rather than raw data.
4. Cross-domain applications using real-time data from multiple sources can be seamlessly implemented.

The current state of art of M2M platforms is quite fragmented and there isn't a single view toward an interoperable smart object world. The M2M commercial platforms are vertically focused on solving specific sector issues and are tightly integrated with applications. This approach, taken from the telemetry legacy applications, has created a bunch of sensor devices not interoperable one each other, with high boundaries and integration possible only at database or presentation layers.

For many years, M2M deployments were based on proprietary/custom applications and networking infrastructures, which were typically expensive to build, manage and maintain. Today's market for sensor devices is a hotbed of idea generation, as the prospect of embedding intelligence in the form of M2M technology into mobile devices has everyone excited about the possibilities. The current market is already filled up with devices that can track everything from blood-glucose levels to traffic patterns.

## 5 Conclusion

The paper discussed a number of popular ICT paradigms, including Cloud computing, IoT and Big Data. It provided an extensive state of the art review of them and the convergence between them. Next, we proposed a M2M system based on a decentralized cloud architecture, general systems and Remote Telemetry Units (RTUs) for E-Health applications. The system was built for Big Data processing of sensors information in the way that data can be aggregated to generate “virtual” sensors, and some measurement results were presented.

With the advent of smart mobile devices, the Internet access has become ubiquitous, and has opened the way for new applications that use M2M communications. Indeed, we envision our work on the Internet of Things paradigm, as an evolution of the traditional Internet for seamlessly integrating most things around us and collecting big data from sensors that track everything happening in our environment.

**Acknowledgments.** This work was supported in part by the EU FP7 ICT Project eWALL project and has been funded by the Sectoral Operational Programme Human Resources Development 2007-2013 of the Ministry of European Funds through the Financial Agreement POSDRU/159/1.5/S/134398.

## References

1. Kahn, E.: Natural language processing, big data, bioinformatics and biology. *International Journal of Biology and Biomedical Engineering* 8, 107–117 (2014)
2. Ochian, A., Suciu, G., Fratu, O., Suciu, V.: Big data search for environmental telemetry. In: *IEEE International Black Sea Conference on Communications and Networking (BlackSeaCom)*, pp. 182–184 (2014)
3. Vermesan, O., Friess, P., Guillemi, P., Gusmeroli, S.: *Internet of Things Strategic Research Agenda*. In: *Internet of Things – Global Technological and Societal Trends*. River Publishers (2011)
4. Suciu, G., Halunga, S., Fratu, O., Vasilescu, A., Suciu, V.: Study for Renewable Energy Telemetry using a Decentralized Cloud M2M System. In: *IEEE 15th International Symposium on Wireless Personal Multimedia Communications (WPMC)*, pp. 1–5 (2013)
5. McFedries, P.: The cloud is the computer. *IEEE Spectrum* 45(8), 20–22 (2008)
6. Hassan, M.M., Song, B., Huh, E.N.: A framework of sensor-cloud integration opportunities and challenges. In: *Proceedings of International Conference Ubiquitous Information Management Communication*, pp. 618–626 (2009)
7. Fox, G.C., Kamburugamuve, S., Hartman, R.D.: Architecture and measured characteristics of a cloud based internet of things. In: *International Conference on IEEE Collaboration Technologies and Systems (CTS)*, pp. 6–12 (2012)
8. Kranz, M., Holleis, P., Schmidt, A.: Embedded Interaction - Interacting with the Internet of Things. *IEEE Internet Computing* 14(2), 46–53 (2010)
9. Jara, A.J., Genoud, D., Bocchi, Y.: Sensors data fusion for Smart Cities with KNIME - A real experience in the SmartSantander Testbed. In: *2014 IEEE World Forum on Internet of Things (WF-IoT)*, pp. 173–174 (2014)
10. McAfee, A., Brynjolfsson, E., Davenport, T.H., Patil, D.J., Barton, D.: Big Data. The management revolution. *Harvard Bus Rev* 90(10), 61–67 (2012)
11. The 3rd Generation Partnership Project (3GPP), TS 23.888 “System improvements for Machine-Type Communications (MTC),” Version 11.0.0 (2012)
12. Saad, W., Abbes, H., Jemni, M., Cerin, C.: Designing and implementing a cloud-hosted SaaS for data movement and sharing with SlapOS. *International Journal of Big Data Intelligence* 1(2), 18–35 (2014)
13. Shah, T., Rabhi, F., Ray, P.: Investigating an ontology-based approach for Big Data analysis of inter-dependent medical and oral health conditions. In: *Cluster Computing*, pp. 1–17 (2014)
14. Eckstein, R.: *Interactive search processes in complex work situations - a retrieval framework*. University of Bamberg Press 10, 62–67 (2011)