

Big Data Analytics in Intensive Care Units: challenges and applicability in an Argentinian Hospital

Javier Balladini, Claudia Rozas
Computer Engineering Department - National University of Comahue
Neuquén, Argentina

{javier.balladini, claudia.rozas}@fi.uncoma.edu.ar
and

Emmanuel Frati
National University of Chilecito
La Rioja, Argentina
fefrati@undec.edu.ar

and
Néstor Vicente, Cristina Orlandi
Lopez Lima Hospital
Río Negro, Argentina

nestorvicente07@yahoo.com.ar, orlandi.mariacristina@gmail.com Francisco

ABSTRACT

In a typical intensive care unit of a healthcare facilities, many sensors are connected to patients to measure high frequency physiological data. Currently, measurements are registered from time to time, possibly every hour. With this data lost, we are losing many opportunities to discover new patterns in vital signs that could lead to earlier detection of pathologies. The early detection of pathologies gives physicians the ability to plan and begin treatments sooner or potentially stop the progression of a condition, possibly reducing mortality and costs. The data generated by medical equipment are a Big Data problem with near real-time restrictions for processing medical algorithms designed to predict pathologies. This type of system is known as real-time big data analytics systems. This paper analyses if proposed system architectures can be applied in the Francisco Lopez Lima Hospital (FLLH), an Argentinian hospital with relatively high financial constraints. Taking into account this limitation, we describe a possible architectural approach for the FLLH, a mix of a local computing system at FLLH and a public cloud computing platform. We believe this work may be useful to promote the research and development of such systems in intensive care units of hospitals with similar characteristics to the FLLH.

keywords: Clinical Decision Support System, Real Time, Big Data, Cloud Computing, Intensive Care Units.

1. INTRODUCTION

In a typical Intensive Care Unit (ICU) at Hospital, many sensors are connected to patients to measure physiological data. Commonly physiological data are the vital signs, for example: heart rate, respiratory rate, body temperature, blood oxygen saturation, and systolic and diastolic blood pressure. The vital signs are shown as waves and numbers by monitors at every bedside. These devices issue audible and visual alerts when measures breach standardized-population-based thresholds, indicating a significant risk to the health of patients. When this happens, it is up to physicians and nurses to rapidly process and analyse all this information in order to make medical decisions.

The data generated by medical equipment in intensive care units are a Big Data problem [1, 2]. For example, heart activity monitored by electrocardiography (ECG) can sample 1,000 or more readings a second to construct a waveform that shows the functioning of the heart [3]. That is, 86.4 million of samples a day per patient. Currently, the nurses register measurements in a paper or computer form from time to time, possibly every hour. So, one sample of each vital sign is saved by hour and the others are lost. With this data lost, we are losing many opportunities to discover new patterns in vital signs that could lead to earlier detection of pathologies. The early detection of pathologies gives physicians the ability to plan and begin treatments sooner or potentially stop the progression of a condition, possibly reducing mortality and costs.

Nowadays, the current state of technology makes it possible to think in systems to process, in real-time [4], the large amount of patient's physiological data, predicting early pathologies by running predesigned and validated algorithms for detecting patterns based on historical cases. In this type of system, different predictive algorithms to detect sepsis or apnea, such as [5, 6], can be implemented. This new system may be able to identify patterns in physiological data and quickly alerts clinicians to danger signs in patients. A system of this type should provide support to extracting knowledge from historical physiological data, coming from multiple health institutions, and generates new algorithms for early detection of new diseases. This type of system is known as real-time big data analytics systems. There are few studies in the literature about this type of systems oriented to ICUs, and between of them we can cite [2, 7, 8].

Located in the city of General Roca, Río Negro, Argentina, the Francisco Lopez Lima Hospital (FLLH) is the most advanced public hospital of the province, but with relatively high financial constraints. It has two ICUs, one for neonatal patients and another for adults, both with the same characteristics as the typical ICUs described above. We have proposed to study the applicability of a real-time big data analytics system for both ICUs. In this paper, we describe the current state of the intensive care units at FLLH detailing the used medical equipment and the possibilities to ex-

tract physiological data from the equipment, and the solutions proposed in the bibliography about real-time big data analytics systems for ICUs and its applicability in the ICUs at FLLH. The novelty of this paper is in the study of applicability and adoption of existing system architectures in a public hospital with relatively high financial constraints.

The rest of this paper is organised as follows. Section 2 describes the generalities of the FLLH, the organization of their ICUs and the work that these units perform, and the medical equipment available in these units. Also describes the big data problem that involves the physiological data. Section 3 shows the characteristics of existing system solutions. Section 4 discusses the applicability and adoption of existing systems by the FLLH. Finally, section 5 presents the conclusions and future works.

2. FRANCISCO LOPEZ LIMA HOSPITAL

Francisco Lopez Lima Hospital (FLLH) is located at General Roca, province of Río Negro, Argentina. The hospital has services of: inpatient evaluated by clinical physicians, surgery, paediatrics, tocogynecology, and auxiliary services of diagnostic and treatment. These services correspond to the category of complexity VI certified by the Argentine Ministry of Health. The hospital has 145 beds.

The Adult Intensive Care Unit (AICU) at FLLH provides comprehensive and continuous care for adult persons who are critically ill and who can benefit from treatment, providing a good die for unrecoverable patients. This unit was inaugurated in 1997 and currently has 7 multipurpose beds. The Neonatology Service at FLLH was opened in 1988 and consists of the following units and capacities: Neonatal Intensive Care Unit (NICU) – 6 beds, Neonatal Intermediate Care Unit – 6 beds, Neonatal Multipurpose Isolation Unit – 2 beds, Pre-discharge Inpatient Unit – 4 beds. NICU is an intensive care unit specializing in the care of ill or premature newborn infants.

The AICU has 6 intensivists, 1 cardiologist, 2 intensive care resident physicians, 12 nurses, and 2 bioengineers. In the Service of Neonatology there are 7 neonatologists, two of them are in the NICU, 1 neonatologists on call, and 1 neonatologist resident physician. There are 18 nurses, 3 of them are in the NICU. There are 2 physical therapists for all units at hospital (i.e., not exclusive to the AICU and NICU). Many services required by the AICU and NICU are provided by the own hospital, such as: X-rays, ultrasounds, laboratory tests, and consultations with specialists. Other services are requested to other medical institutions as: tomography, catheterization and magnetic resonance imaging.

According to the latest AICU's statistics, the number of annual patients is 374, with 17.6% of mortality, 88.2% of occupancy, 6.2 patients per day, and an average stay per patient of 5.6 days. The hospital produces statistics of the neonatology service, but the data are not broken down by each unit, including the NICU. The most common diseases in the AICU are traumatic brain injury and septic shock. In the NICU, the most common diseases are the preterm birth, respiratory distress and sepsis. All statistics are calculated by a statistical application software provided by the Argentine Society of Intensive Care.

Regarding to Internet connectivity, the FLLH has no Internet leased line, instead they use shared Internet access lines. A leased line provides a dedicated connection be-

tween the Internet Service Provider (ISP) and the contracting institution, and includes a service level agreement that defines performance, reliability and support criteria. Shared accesses are cheaper because they are shared by multiple users. However, the quality of service is not guaranteed and, when the number of users in the area increases, the performance is reduced. Different offices at FLLH use Internet shared access through ADSL (Asymmetric Digital Subscriber Line) or Cable services.

Medical data and equipment

We refer to medical data as all data related to the health of a patient. For our purpose, we need to classify the medical data according to the recording continuity of data. A medical dictionary defines “continuous”¹ as “continuing without intermission or recurring regularly after minute interruptions”. So, we classify the data as continuous or noncontinuous data. In this sense, for example, clinical data, medical images or laboratory data are considered as noncontinuous data, and pulse rate or temperature is considered as continuous data. We are focused on continuous data because this is the origin of the problems addressed by the real-time big data analytics systems.

The medical equipment is designed to aid in the diagnosis, monitoring or treatment of medical conditions. The rest of this section describes the different medical equipment available in the intensive care unit both adult and neonatal of FLLH, analyzing the data output capabilities of each device. We can classify the available medical equipment in:

- Monitoring equipment used to measure the medical state of patients: intracranial pressure monitors, and medical monitors that show vital signs (body's basic functions) and other parameters.
- Environmental maintenance equipment: infant radiant warmers and neonatal incubators.
- Life support equipment used to maintain a patient's bodily function: medical ventilators.
- Treatment equipment: infusion pumps and sequential compression devices.
- Diagnostic equipment: medical imaging machines, electroencephalograph.
- Laboratory equipment: automates or helps analyse blood, urine, and dissolved gases in the blood.

As available diagnostic and laboratory equipment produces noncontinuous data (at least currently), we are not worried about this type of equipment and its data. The remaining types of equipment include devices that can produce continuous data.

Some devices are more used than others, according to the medical complexity of each patient. The medical equipment most used is the medical monitor because all patients, from low to high medical complexity, have connected one of these. A medical monitor is capable of measure many parameters, commonly: electrocardiography (ECG), heart rate (HR), noninvasive blood pressure (NIBP), invasive blood pressure (IBP), functional arterial oxygen saturation (SpO₂), pulse rate (PR), respiration rate (RR), capnography (EtCO₂ and InCO₂), and temperature (Temp). Table

¹<http://www.merriam-webster.com/dictionary/continuous>



Figure 1: Baby in an incubator and a medical monitor at the NICU of FLLH

1 and 5 show the available medical monitors at NICU and AICU, respectively.

In the NICU, there is other equipment that each baby need: an infant radiant warmer or a neonatal incubator. This equipment is used to maintain environmental conditions suitable for baby. Neonatal incubators are used for preterm babies or for some ill full-term babies, and infant radiant warmers are used for full-term babies. Both equipments consider the body temperature to set an appropriate environment for the baby. Table 4 shows the available incubators and infant radiant warmers at NICU, and Figure 1 shows a photograph of a baby in an incubator and the medical monitor at the NICU.

At higher complexities, intracranial pressure monitors, infusion pumps, sequential compression devices and medical ventilators are connected to patients. The intracranial pressure (ICP) monitoring is used in treating severe traumatic brain injury of adult patients. All current clinical available measurement methods are invasive and the most used need the insertion of a catheter into the cranium. Table 8 shows the available ICP monitors at AICU. An infusion pump infuses fluids, medication or nutrients into a patient’s circulatory system. An infusion pump is used to administer fluids in ways that would be impractically expensive or unreliable if performed manually by nurses. For example, they can administer as little as 0.1 millilitre per injections (too small for a drip), injections every minute, or fluids whose volumes vary depending on the time of day. Table 3 and 7 show the available infusion pumps at NICU and AICU, respectively. A sequential compression device (SCD) is designed to im-

Count	Brand	Model	Connectivity
5	BLT	V6	Ethernet, USB, RS-232
1	Massimo	Radical RDS-3	Serial RS-232 and analog output

Table 1: Medical monitors at NICU

Count	Brand	Model	Connectivity
3	Neumovent	Graph	RS-232
1	Medix	SLE 5000	RS-232

Table 2: Ventilators at NICU

prove venous circulation in the limbs of adult (not neonates) patients. This type of device does not produce useful data and, therefore, they have no communications ports. A medical ventilator is a machine designed to mechanically move breathable air into and out of the lungs, used when a patient is physically unable to breathe, or breathing insufficiently. Table 2 and 6 show the available medical ventilators at NICU and AICU, respectively.

There is a mix of old and new equipment, but almost all have some class of analog and/or digital output. Most systems have RS-232 (Recommended Standard 232) serial connection. RS-232 is a serial port, that is, a physical interface through which information transfers in or out one bit at a time. Normally, modern computers have no serial ports and require serial to Universal Serial Bus (USB) or serial to Ethernet converters to allow compatibility with RS-232 serial devices. A few systems have analog output, and this output need to be converted to digital signals in order to get data from a computer. This conversion is not easy and require the use of an additional device, for example, a USB oscilloscope. Only one monitoring system has Ethernet and USB connections, that simplify the transmission of data from the monitoring system to a computer. The major problem is that the manufacturers normally do not describe how to get medical data from communication ports on systems.

Another problem arises because many devices of the same type are developed by different manufacturers. They are seven clinical monitors of four different brands at AICU, preventing the use of a conventional remote telemetry monitor (a central monitor that allow to see the measurements of all patients). This situation occurs due to low financial resources which force to only replace old and broken devices.

The real-time big data problem

As we said in the previous subsection, every patient at AICU or NICU is connected to a medical monitor. The medical monitors measure different parameters at different frequencies, and some parameters need highest frequency sampling than others in order to achieve high accuracy measurements. For example, temperature is a parameter that requires low sampling rates, because the temperature changes very slowly. The volume of data generated by the measurements of a parameter increases as the sampling frequency increases. Considering all parameters commonly monitored

Count	Brand	Model	Connectivity
9	Volumat	Agilia	8 pin serial communication port
2	Inyectomat	MC Agilia	Not available
1	Adox	Activia	Not available
1	Atom	Syringe PVM 1235N	Not available

Table 3: Infusion pumps at NICU

Count	Type	Brand	Model	Connectivity
2	incubator	Medix	Natal Care LX	Not available
5	incubator	Medix	PC-305	Not available
2	radiant warmer	Medix	SM-401	Not available

Table 4: Incubators and radiant warmers at NICU

Count	Brand	Model	Connectivity
3	Feas	8272 and 6628	RS-232
2	Nihon Kohden	Life Scope LC, BSM-3101	Analog output (only electrocardiography and blood pressure) and proprietary (ZB) socket
1	BCI	Advisor 9200	RS-232 and analog output
1	Mediana	Lucon M30 (M30M-3LNS)	Ethernet and USB ports

Table 5: Medical monitors at AICU

Count	Brand	Model	Connectivity
3	Neuomvent	Graph	RS-232
2	Siemens	Servo Ventilator 300/300A	RS-232
2	Taema	Osiris	Not available
1	Neuomvent	Cirrus	Not available
1	Dräger	Evita	RS-232

Table 6: Ventilators at AICU

Count	Brand	Model	Connectivity
5	Hospira	PlumA+	RS-232
4	Volumat	Agilia	8 pin serial communication port
3	Medifusión	MP-1000	Not available
2	Baxter	Flo-Gard 6201	RS-232

Table 7: Infusion pumps at AICU

Count	Brand	Model	Connectivity
1	Camino	V420-6	Not available

Table 8: Intracranial pressure monitors at AICU

by these devices, the sampling frequency of the ECG parameter is one of the biggest producers of data.

According to Mediana Lucon M30 Operator's manual (a medical monitor at AICU), the monitoring of ECG parameter can detect pacer pulses as little as 0.1msec of width. So, the parameter' reading frequency need to be of 1000 samples per second. This large amount of data generated per unit time produces a high computing and storage demand to be supported by the analytic system under study. The computing demand is produced by the near real-time execution of algorithms designed for early prediction of pathologies. These algorithms process large amounts of physiological data and need to be executed for each patient for which it were designed; for example: adult or preterm neonate in a specific week. Furthermore, a high demand of secondary storage is required to maintain all historical physiological data of all patients. The purpose of having historical data is the extraction of knowledge from them, in a form of behavior patterns that allow to predict a certain pathology, to later use this knowledge in the development of new predictive algorithms. Furthermore, historical data is required to validate algorithms, possibly some part of them are used for knowledge extraction and another part for validation. Considering 1000 samples per second, it is translated to 86.4 million readings a day per patient. The representation of each point as an IEEE-754 single precision floating point

number need 4 bytes. Therefore, the volume of data is approximately of 330 MiB a day per patient, that is, a total of 2,25 GiB a day for all 7 beds at AICU. However, this may even be worse. Newer converters can sample at 10000 to 15000 per second or even higher [3] (other converters are adaptive in sampling rate, with output that is proportional to the energy detected). So, the volume of ECG data can be very much bigger than the calculated.

Currently, all measurements obtained by the medical equipment at AICU and NICU are lost. Sometimes, the physicians need to read previous data, but it is impossible if they do not explicitly program the machines to save measurements. But, the machines can only store a few minutes in their internal memories.

3. EXISTING SOLUTIONS

A clinical decision support system (CDSS) is a health information technology system that is designed to assist health professionals with clinical decision-making tasks. Most of the CDSS systems work with non continuous data (see Section Medical data and equipment), i.e., the data normally used by them are saved on clinical stories as: clinical data, diagnostic test and other data which could provide relevant information for detection of diseases. To the best of our knowledge, only one CDSS system has been proposed for continuous data: the Artemis system, a real-time big data analytics system for ICUs.

Since early 2009, the Hospital for Sick Children, also known as SickKids, the University of Ontario Institute of Technology (UOIT) in Oshawa, Ontario, Canada, and IBM collaboratively designed, built and deployed Artemis in SickKids NICU. The Artemis Project research team is headed by Dr. Carolyn McGregor, UOIT's Canada Research Chair in Health Informatics. The team is composed neonatologists, emergency physicians, nursing staff, computer scientists and engineers. The result of the project is an online health analytics platform that support the acquisition and storage of physiological data streams and clinical information system data of multiple patients with the objective of analysing this data in real-time for earlier disease detection. The Artemis Project's success at SickKids has led other hospitals to join the study: two hospitals in China (in Shanghai and Shenzhen), one in Australia and one in the USA.

Section Toronto's SickKids Hospital describes the SickKids Hospital because it is the medical institution directly involved in the project. Section Artemis presents the first implementation of Artemis described in [2, 7]. Finally, Section Artemis Cloud presents the second implementation of Artemis, the Artemist Cloud, described in [8].

Toronto's SickKids Hospital

SickKids is the major paediatric hospital in Toronto, Ontario, Canada. SickKids has built an integrated environment of patient care, research and learning for to improve health of children in the country. The hospital has actually twelve centres to formalize and enhance the collaborations between researchers and clinicians. These centres are specialized in bone health, brain and mental health, healthy active kids, cystic fibrosis, cancer, genetic medicine, global child health, heart, image-guide care, heart, pain, transplant and regenerative medicine, and inflammatory bowel disease.

The SickKids NICU has 36 beds and provides services for

newborns up to four weeks of age. Annually, the NICU admits approximately 800 neonates with common conditions which include prematurity, acute respiratory illness, neurological disorders, surgical emergencies, cardiac disease, and genetic and metabolic disorders. The NICU utilizes ECG and derived signals such as heart rate, respiration rate, and chest impedance for breath detection together with systolic, diastolic, and mean blood pressure data to monitor the status of patients.

Artemis

The Figure 2 shows the general architecture of the first Artemis platform. At SickKids is located the data acquisition and online analysis components of the platform while the data persistency, knowledge extraction, and redeployment components reside at UOIT. Artemis supports clinical studies on late-onset neonatal sepsis, apnea of prematurity, retinopathy of prematurity and pain.

The platform has five components. The “data acquisition” component obtains physiological data streams from medical devices and available clinical information. The interface between the wide range of medical devices used in an intensive care environment and Artemis is done with a set of hardware and software elements (provided by Capsule Tech Inc). The patient monitors are connected to a DataCaptor terminal unit which converts the RS-232 output to an Internet Protocol (IP) stream over Ethernet. Next these data are forwarded to a Capsule DataCaptor Interface Server that can support up to 500 simultaneously connected devices. The server has the function to filter the data received, to extract data streams necessary for the study, to convert these streams to Artemis format, and to send them to the Medical Data Hub. This hub receives data from the server and create concurrent data streams for the streaming system. Furthermore, the CIS Adapter gets patient data from the clinical information system (CIS) and streams the data to the Artemis clinical rules. A clinical rule is a medical algorithm that can be defined by a clinician or can be proposed through the extraction of knowledge from physiological data streams, laboratory results, and observations of a patient.

The “online analysis” component employs an InfoSphere Streams middleware system of IBM to process data in real-time applying the appropriate algorithm. The characteristics of the patient determine the number of algorithms to be applied and for how long. This component provides scalable processing of multiple streams. The programming of the stream computing system is done with the Stream Processing Application Declarative Engine (SPADE) language [9]. SPADE is a high-level declarative language for high-performance distributed stream processing systems that permit flexible composition of parallel and distributed data-flow graphs.

The original data and generated analytics are stored in the “data persistency” component. The Data Integration Manager (DIM) has a set of SPADE operators which interact with an open database system. The DIM receives the data and stores it in an appropriate database. Periodically, the Data Mover moves the data stored in the database, formatted for data mining, to the repository in the knowledge extraction component.

The “knowledge extraction” component uses a data stream data mining framework for multidimensional real-time data analysis. The output of this component is a clinical rule and data specifically tailored to enable clinicians to perform

clinical research in a range of condition.

The “redeployment” component translates new acquired knowledge into a SPADE representation and feeds new clinically validated algorithms to the online analysis component.

In regard to the data privacy, as the collected data are personal healthcare data, the collected and stored data in Artemis do not directly identify infants. When an infant is enrolled, an Artemis identifier is generated. Furthermore, the medical monitor has a unique identifier. The data transmitted from the medical monitor are associated with this unique identifier. The hospital maintains the association between a bed and the medical monitor, and the association between the medical monitor identifier and the Artemis identifier is placed in a mapping database table.

With regard to data safety, the data are transmitted on the hospital network that has been secured. The connection between the computers at SickKids and UOIT is made through a secure tunnel that uses the SSH2 protocol encrypted with a 4096-bit RSA key. The processing components and Artemis database system are in a safe place accessible only to Artemis team members. To control operations that users can perform on the database system, the built-in authentication system of the operating system and the authorization mechanism and access control system of the database system are used. Also, the system is separate from all other hospital systems. The UOIT performs a mirror copy of the Artemis database on a dedicated computer located in a secure locked room, only accessible to Artemis team members. The IBM team maintains a mirror from the UOIT mirror, using the same secure tunnel mechanism used between SickKids and UOIT. This mirror is stored on a protected machine with password in a secure laboratory, and only the IBM team can access the data.

Artemis Cloud

The first Artemis platform assumes that the Data Acquisition and Online Analysis components are located in the same location as the medical devices that provide the data. However, this system represents a negative cost model for rural or remote small hospitals. For hospitals in this situation, Artemis Cloud is a cloud computing platform that enables the use of computing resources outside of the healthcare facility, reducing costs of technical support staff and required infrastructure, and allowing patients access to better care without the need to refer them to an urban health centre.

Artemis Cloud is based on the provision of software through the Software-as-a-Service and the provision of the storage of persistent data through the Data-as-a-Service model. In the Artemis Cloud, the interface with Artemis is performed through of web services that allows interaction of hospitals with the system. Also, the hospitals have access to persistent data stored in the cloud. The Figure 3 shows the framework.

The Artemis Cloud has been deployed on servers located in a secure server room at UOIT. The Artemis Cloud receives data from hospitals in HL7 format, which is the set of international standards for transfer of clinical and administrative data between software applications.

Still there are no health statistical results showing the impact of Artemis in the Hospital.

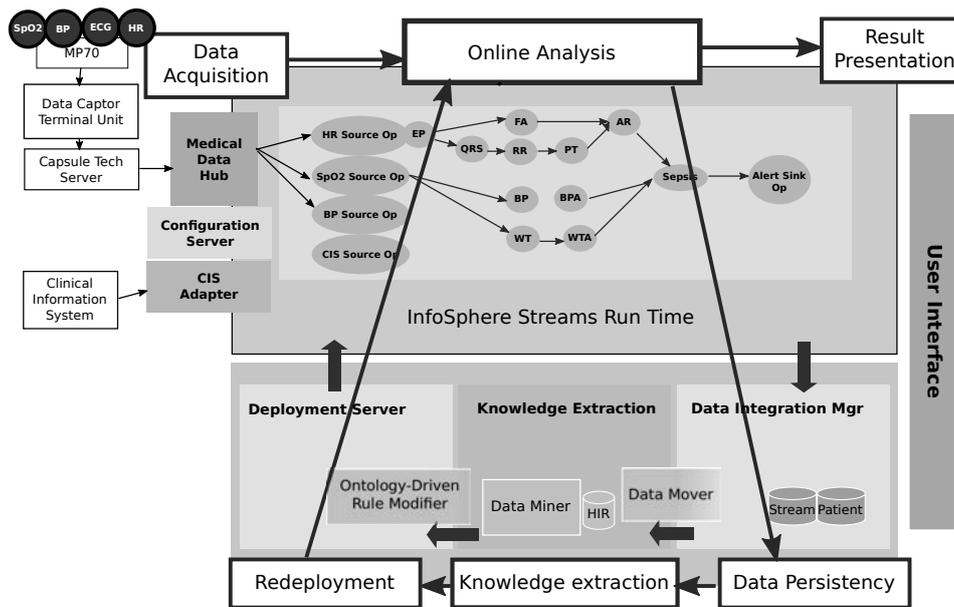


Figure 2: Artemis platform

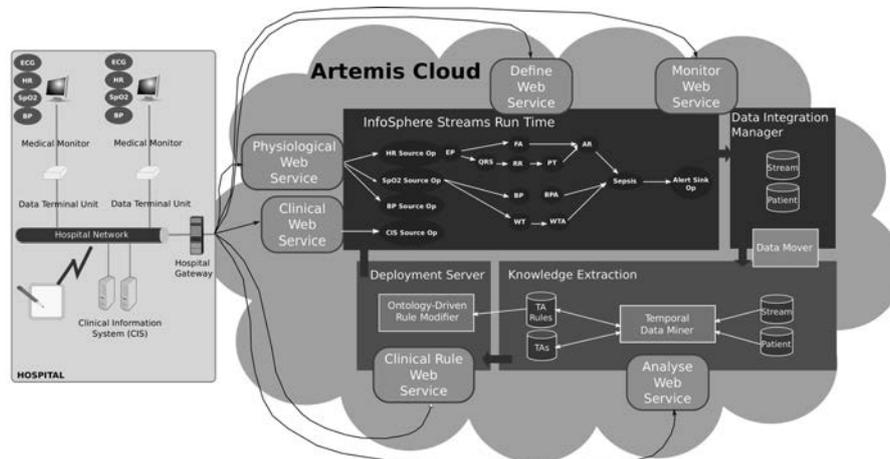


Figure 3: Artemis Cloud Framework

4. CAN EXISTING SOLUTIONS BE APPLIED IN FLLH?

Artemis and Artemis Cloud have Data Persistency and Knowledge Extraction components located in remote servers to the hospital (servers placed at the university which develop the system), and we believe that this approach is not appropriate. The main problem is that we do not have a private computing system for this purpose, and we believe that a public cloud is more appropriate when data storage and computing need to be highly scalable. A public cloud will support the increase of the number of hospitals.

Furthermore, in Artemis Cloud, the Online Analysis (of medical data) component is performed in remote servers to the hospital. In this way, in order to process the high rate physiological data, they need to be transmitted from the hospital towards the remote site before. When some pattern is detected, an alert is transmitted back to the hospital. Considering the unreliable Internet access of FLLH through shared lines (cable and ADSL), this approach is not suitable to be applied in FLLH mainly because in the presence of a

pathology, there are no guarantees about the medical alerts will arrive to the hospital in near real-time (from the moment the physiological data of the medical equipment were acquired). Moreover, the Internet access could be interrupted not only by a few minutes or hours, but several days, because when the service is interrupted the local ISP can normally delay between 3 to 4 days to visit the client and, in consequence, to restore the service. The reason why the hospital does not have an Internet leased line is unknown.

In Figure 4 we show a possible architectural approach of a real time big data analytics system suitable for the FLLH. In this system, the online analysis is performed on a dedicated server placed inside the hospital. The system and the medical algorithms implementation must offer a high performance solution with relatively cheap acquisition cost, in order to support near real-time computing. The system has to be scalable to support the growing amount of inpatient, and high energy efficiency to reduce the maintenance cost (and possibly the installation cost of high power supply lines). The inpatient data are saved temporarily in the hospi-

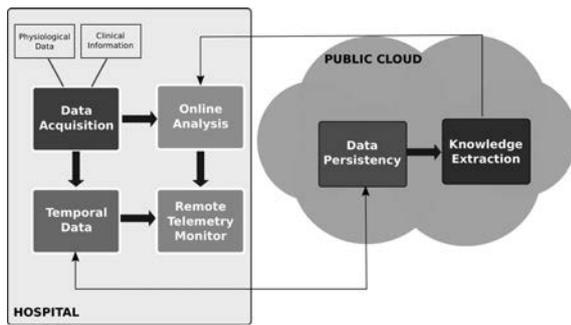


Figure 4: A possible architectural approach suitable for the FLLH

tal in order to permit the remote telemetry monitoring with real-time data and historical data. The telemetry monitors are also used to show alerts of pathologies prediction.

The acquired inpatient data, from any hospital, are saved in a persistent way in a public cloud. It must be observed that the temporal storage on hospitals needs to save only data of current inpatients. This means the storage does not need to scale simply by the passing time, but only when grows the amount of beds of the ICUs. Instead, the persisting storage requires to scaling with the passing time because it must save the new medical data that are received from the hospitals. When a patient arrives to an ICU, if previous data exist, they can be transmitted from the persistent storage to the temporal storage of the corresponding hospital. The resources of the cloud will be used in order to perform an knowledge extraction of the medical data persistently saved and to generate new medical algorithms for pathologies early detection. After, these medical algorithms will be implemented and will be distributed between the online analysis components of the hospitals. The knowledge extraction will be performed through a system located in the public cloud, accessible through a web service, which will provide the necessary tools for that physicians to discover new patterns of pathologies in early way.

5. CONCLUSIONS AND FUTURE WORKS

This paper has presented the current state of the adult and neonatal intensive care units at Francisco Lopez Lima Hospital (FLLH). We have detailed the medical equipment used and the possibilities to extract physiological data from them. We have found it is not easy to extract data from the medical equipment to a computer because this type of information is not included in the default documentation provided by the manufacturers with the machines. Also, we provide a brief survey on the solutions proposed in the literature about real-time big data analytics systems in intensive care units (ICU). Finally, we have studied the applicability of these solutions in the ICUs at FLLH. Respect to the applicability of existing solutions of real-time big data analytics systems at FLLH, we have found that they are not suitable. One approach need an Internet leased line and it can not be used with shared Internet access lines. The problem lies in the impossibility to send big physiological data from the hospital to the private cloud computing platform and receive the medical findings in near real-time using unreliable communications. Both analysed approaches use private servers for highly scalable storage and computing, but we believe that a public cloud is more appropriate for this purpose supporting a greater increase of the number of hospitals and beds. Taking into account these limitations, we describe a possi-

ble architectural approach needed to develop an appropriate real-time big data analytics system for the FLLH, a mix of a local computing system at FLLH and a public cloud computing platform. We believe this work may be useful to promote the research and development of such systems in intensive care units of hospitals with similar characteristics to the FLLH.

6. ACKNOWLEDGEMENTS

We thank to the intensivist Romina Villegas and the neonatologist Diego Acosta of the Francisco Lopez Lima Hospital for your collaboration to provide essential data to the research summarised here.

REFERENCES

- [1] A. Gandomi and M. Haider, "Beyond the hype: Big data concepts, methods, and analytics," *Int J. Information Management*, vol. 35, no. 2, pp. 137–144, 2015.
- [2] C. McGregor, "Big data in neonatal intensive care," *Computer*, vol. 46, no. 6, pp. 54–59, 2013.
- [3] P. Kligfield, L. S. Gettes, J. J. Bailey, R. Childers, B. J. Deal, E. W. Hancock, G. van Herpen, J. A. Kors, P. Macfarlane, D. M. Mirvis, O. Pahlm, P. Rautaharju, and G. S. Wagner, "Recommendations for the Standardization and Interpretation of the Electrocardiogram: Part I: The Electrocardiogram and Its Technology: A Scientific Statement From the American Heart Association Electrocardiography and Arrhythmias Committee, Council on Clinical Cardiology; the American College of Cardiology Foundation; and the Heart Rhythm Society Endorsed by the International Society for Computerized Electrocardiology," *Circulation*, vol. 115, no. 10, pp. 1306–1324, 2007.
- [4] M. Barlow, *Real-Time Big Data Analytics: Emerging Architecture*. O'Reilly, Jun. 2013.
- [5] K. D. Fairchild, "Predictive monitoring for early detection of sepsis in neonatal icu patients," *Curr Opin Pediatr*, 2013.
- [6] H. Lee, C. G. Rusin, D. E. Lake, M. T. Clark, L. Guin, T. J. Smoot, A. O. Paget-Brown, B. D. Vergales, J. Kattwinkel, J. R. Moorman, and J. B. Delos, "A new algorithm for detecting central apnea in neonates," *Physiological Measurement*, vol. 33, no. 1, p. 1, 2012.
- [7] M. Blount, M. Ebling, J. Eklund, A. James, C. McGregor, N. Percival, K. Smith, and D. Sow, "Real-time analysis for intensive care: Development and deployment of the artemis analytic system," *Engineering in Medicine and Biology Magazine, IEEE*, vol. 29, no. 2, pp. 110–118, March 2010.
- [8] C. McGregor, "A cloud computing framework for real-time rural and remote service of critical care," in *Computer-Based Medical Systems (CBMS), 2011 24th International Symposium on*, June 2011, pp. 1–6.
- [9] B. Gedik, H. Andrade, K.-L. Wu, P. S. Yu, and M. Doo, "Spade: The system s declarative stream processing engine," in *Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data*, ser. SIGMOD '08. New York, NY, USA: ACM, 2008, pp. 1123–1134.