Criteria of Evaluation for Systems Using Sensor as a Service

A discussion about repositories and search engines on S²aaS systems

Anderson Brasiliense de Oliveira Brito
Diretoria de Tecnologia da Informação
Instituto Federal do Amapá, IFAP
Macapá, Brasil
Email: anderson@ifap.edu.br

Felipe Silva Ferraz
Centro de Informática
Universidade Federal de Pernambuco, UFPE
Recife, Brasil
Email: fsf3@cin.ufpe.br

***Abstract*—This document proposes an approach to the criteria settings for evaluation of systems that use sensors as a service through the analysis of sensor repository and search engine data. The sensor paradigm as a service is a branched concept of cloud computing, which is still evolving. Thus, to achieve the end user's expectation for this type of service, it is necessary to define clear parameters for the evaluation of the delivery of it. The proposal presented in this paper provides an analysis based on the type of sensors and how its owner groups them in a multiuser system.**

Keywords- sensor as a service; sensors repository; search engine data.

I. INTRODUCTION

The society has seen an expansion of emerging technologies for the Internet. This phenomenon is a favorable environment for several factors, among which are: the emergence of IPv6 as a protocol with the possibility of enabling many devices connected to a network, the price of sensors, processors and network devices which have their price declined over time, as well as the wireless network for computers, WiFi, which served as input for connecting devices both at home and in offices.

From this scenario, the market expects an environment called "Industrial Internet" will provide 10-15 trillion dollars in the next 20 years. Based on this forecast, it was created an economic value, called "Internet of Everything" until 2020 [1].

The cloud becomes the most suitable environment to support the great mass of devices that will be connected to this global communication network, with a forecast of 50 to 100 billion of connected devices in the aforementioned period [2], and in this scenario, the type node or sensor devices will account for 60% of the total available on the Internet [3].

The delivery of this information is a challenge, considering that there is no standard interface for sensor communication. Another difficulty would be how to manage repository sensors so that it can be qualified and can provide useful data to a system.

The analysis of a sensor repository would enable decision making by the system based on its business model. In an environment of service to the end user where the supply would be the sensor data, the repository qualification will be crucial, because the expectation in the search data held by the user will depend directly of the sensors that the repository provides.

Considering paradigms such as Smart Cities, where the citizen is inserted in this context with the intention to shape innovation and urban development through their participation [4], delivery of these services, which in part can be supplied from sensor data, should take into consideration the repository where such devices are contained. In this sense, mean repository as sensors virtualization that provides data feeds on one or more systems.

This model proved to be very efficient for the market, because companies would not need to invest part of their capital in IT assets, transferring this responsibility and risks to third parties [5].

Some authors [15] propose a modelling for smart cities, in which four layers are defined, namely of 1) sensors and sensor owners, 2) sensor publishers, 3) extended service providers, and 4) sensor date consumers.

The purpose of this article is to provide the analysis of layers 1 and 2. These layers were used as the basis for the delivery service, because they are the basis of data consumption. Without them, the others would not receive data for end users.

Manzoor [16] proposes criteria for quality context QoC, based on the quality information analysis from data obtained from sensors.

Thus, the approach in the definitions of a sensor repository and the data delivery engine will be discussed through evaluation of sensors.

The remainder of this paper is structured as follows. In Section 2, the cloud computing concepts will be addressed. Section 3 will discuss the criteria as well as analyze the results obtained from the queries generated in the sensor repository. Section 4 concludes the paper.

II. CLOUD AND SENSORS

The Cloud Computing began to be broadcast in October 2007, when IBM and Google decided to establish a partnership to create a new model for computing, based on current characteristics of cloud computing with high availability, computational resiliency, resources on demand, from a high quality system [5].

Many authors [5]-[9] group the cloud computing model into three distinct classes, as follows:

- Infrastructure-as-a-Service (IaaS) - class that is characterized with hardware virtualization on the supply side.
- Platform-as-a-Service (PaaS) - the infrastructure is abstracted from a layer between the hardware and applications through an interface feeding. The Azure and Google App platforms are examples of this class.
- Software-as-a-Service (SaaS) - it aims virtualization of local computer applications to the cloud. This class is the highest level of abstraction, getting under the supplier's responsibility to maintain, update and support from both the hardware and the software, leaving the end user only the service consumption.

However, these classes are subject to change or develop their own concepts, because, according to the National Institute of Standards and Technology (NIST) cloud computing is defined as a paradigm in development and thus their definitions, case use, technologies, risks and benefits will be redefined, based on interactions between the public and private sectors [10]. From this view, other terms have been introduced to the academic community. They include the Sensing and Actuation as a Service (SAaaS), Sensor Event as a Service (SEaaS) sensor as a Service (SenaaS), DataBase as a Service (DBaaS) Data as a Service (DaaS), Ethernet as a Service (AAS), Identity and Plicy Management as a Service (IPMaaS) and Video Suveillance as a Service (VSaaS) [11].

Many of these terms have acronyms identical with other terms and that can be confusing. For example, Image as a Service(IaaS) [12] has the same acronym as Infrastructure as a Service and Sensing as a Service (SaaS) [11] has the same acronym as Software as a Service, which is also represented as S^2aaS [13]. Thus, a broader class could incorporate all other through the Everything as a Service (XaaS) [14].

III. Definition of Analysis Criteria

From an environment where everything can be offered as a service, cloud enables a favorable site for the sensors expansion through their systems virtualization and subsequent delivery to the end user.

A. Analysis about sensor

To be able to define the repository quality criteria, it will be necessary to assess, first, the characteristics of sensors that compose it. Among the listed characteristics to evaluate a sensor, five were selected and used by [16]:

*1) Correction: a*s the sensor ability to measure the actual value close to the real.

The measurement error was proposed by [17], using the equation,

$$E = M - T \tag{1}$$

where E is the measurement error represents by the difference between the actual value T and the value measured by sensor M.

Thus, a physical sensor may have its accuracy value calculated by the equation,

$$C = 1 - \frac{E}{T} \tag{2}$$

where C is the correction value, E the error value and T, the real value. The correction is obtained by subtracting the relative error of 1.

2) Accuracy: the ability of a sensor to provide the same reading on the same measurement on equal terms. Unlike the correction which has a proximity to a true value, the accuracy shows the sensor proximity of successive readings, which can be represented by the equation:

$$P = \frac{true\ positives\ number}{total\ true\ positives\ and\ falses\ number} \tag{3}$$

where P is the accuracy value, true positives number represents the cases that have been correctly recognized as positives, and false positives those that have been incorrectly recognized as positive [16].

3) Time period: is the time interval between two measurements.

4) Sensor State: is related to the environment where the sensor is installed, and can be static (in the case of fixed sensors, such as temperature measurements) or dynamic, in case of integrated sensors on people or mobile device

5) Range: refers to the maximum distance that a sensor can collect a context measurement.

B. Analysis about Repository

The analysis of the repository is performed taking into consideration the sensors quality and the access level of its slices.

Regarding data access level we have the following:

- Public slice: in this type profile, all sensors arranged in this repository will be made available to any user in the system;
- Private slice: in the private slice, the sensor network manager does not allow access to users freely. In this way, only the users created by him or who have requested access by invitation will be able to access the data on these devices;
- Mixed slice: in this profile, the network manager can provide part of the sensors of your slice for any user. The other sensors will not be visible.

Regarding sensor quality we have the following:

The criteria selected to define the sensor quality will be based on correctness and accuracy, since they are objective data and not properly linked to the context.

The sensors will be classified by the sum of correction value with the accuracy value, by equation,

$$Qdt = C + P \tag{4}$$

where Qdt is the sensor quality content, C is the correction value and P the accuracy value.

The total quality of the repository sensors is calculated by equation:

$$QdtRep = \frac{sum\ of\ the\ indices\ for\ each\ sensor}{total\ sensors\ X\ 2} \tag{5}$$

 ISBN: 978-1-61208-438-1

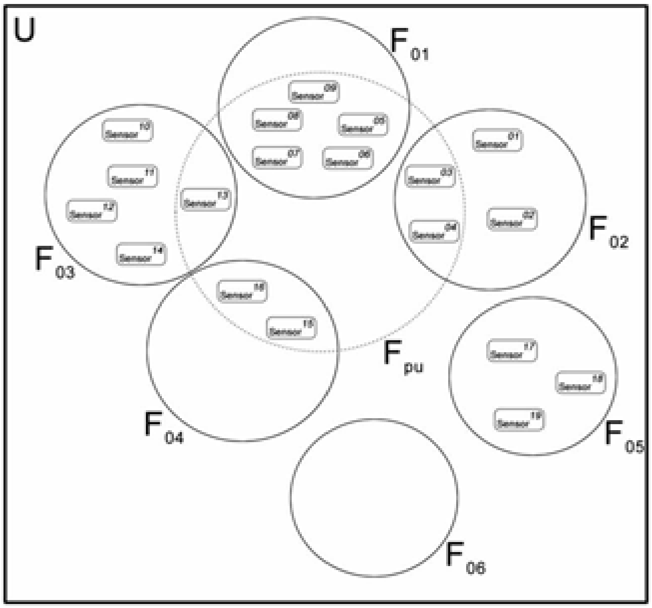

Figure 1. General representation of sensors repository

Below, we find the calculation to assess the repository where Fpu is the public slice and represents total public sensors, being responsible for delivering the information from the sensors contained in this slice for any system user, and U, which represents the total sensors contained in the repository. The universe of sensor system is the sum of all the slices, represented by equation,

$$\text{QftRep} = \frac{\textit{Total public sensors}}{\textit{Total sensors}} \quad (6)$$

where QftRep represents the quality of slices repository.

From these formulas, it is possible to set criteria in order to assess the repository based on the slices and the sensors characteristics, by equation:

$$Qrep = \frac{QdtRep + QftRep}{2} \quad (7)$$

Qrep represents the repository quality through the sums of the sensors qualities and the slices divided by 2. Thus, a higher quality repository is one in which Qrep is closer to 1.

In Figure 1, it is possible to observe the sensors and their respective slices together. Each manager would be responsible for one of the numbered slices. These could be exclusively private (F05), exclusively public (F01 and F04), mixed (F02, F03 and F04) or empty (F06).

For quality of the sensors, will be applied the following values, as shown in Table 1.

TABLE I. TABLE SENSOR CHARACTERISTICS

Item	Access level	Correction	Accuracy	Index
Sensor01	Private	0,3	0,4	0,7
Sensor02	Private	0,9	0,9	1,8
Sensor03	Public	0,5	0,8	1,3
Sensor04	Public	0,9	1,0	1,9
Sensor05	Public	0,5	0,5	1
Sensor06	Public	0,5	0,5	1
Sensor07	Public	0,8	0,8	1,6
Sensor08	Public	1,0	0,8	1,8
Sensor09	Public	1,0	1,0	2
Sensor10	Private	1,0	1,0	2
Sensor11	Private	0,5	0,9	1,4
Sensor12	Private	0,8	0,8	1,6
Sensor13	Public	0,8	0,8	1,6
Sensor14	Private	0,2	0,6	0,8
Sensor15	Public	0,9	0,9	1,8
Sensor16	Public	0,5	0,8	1,3
Sensor17	Private	1,0	1,0	2
Sensor18	Private	1,0	1,0	2
Sensor19	Private	1,0	1,0	2
Total				**29,6**

From this scenario, it is possible to assess this repository and the system can assign value to it.

Applying the formulas for the scenario presented in Figure 1, we have the following data in Table 2:

TABLE II. SCENARIO

Equations Aplication	Index
$\text{QdtRep} = \frac{29{,}6}{38}$	0,78
$\text{QftRep} = \frac{10}{19}$	0,53
$\text{Qrep} = \frac{0{,}78 + 0{,}53}{2}$	0,66

The closer to 1 value, the better the repository quality, from the characteristics of each sensor as the access level assigned to them by means of each slice.

C. Analysis about data search engine

The search engine is the service responsible for the search of sensors available for each user profile. This search is made in slices, where the user has access. By default, any user can receive research data from the primary slice. The search will retrieve data from other slices only if their available sensors are marked as public.

In the survey, the user can enter a sentence with the parameters that are related to desired data. For example, if he wants air humidity data in a particular city, the sentence could be "humidity Sao Paulo." The search interface then looks for the repository data based on this query.

As a proposal for a model of sensor channels, a representation of this channel was implemented, as the class in Figure 2. From this implementation, it is possible to define the search engines on their attributes.

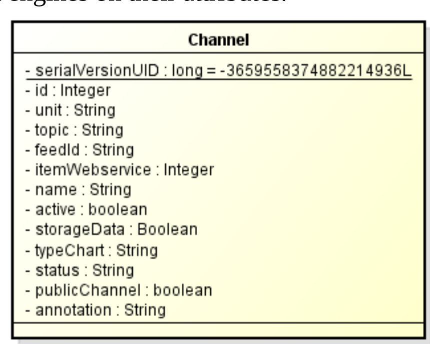

Figure 2. Sensor channel class

Search engines perform an analytical research on the unit, name and annotation attributes of the channel object.

Among the attributes of these objects, these three were chosen because they represent the measurement characteristics. The attribute unit refers to the measurement unit used by the channel, e.g., °C (degree Celsius) for temperature

The name field refers to a measurement identification name such as "TEMP". Finally, the annotation field serves as a comments field about this sensor channel, where the sensor manager could enter comments about it, as "Temperature capture in the Boa Vista neighborhood in Recife".

The analytic research is done by Hirbenate Search library implementation, which is a tool that integrates the Apache Lucene technology of complete search engine for text, with implementation by Index Hibernate by domain from notes, index database synchronization through objects [18].

The implementation of this library is made on the object channel by setting the unit, name and annotation attributes, with the definition of the Ngram type parameters, which is a feature of this API for data analysis, filtering search of words with 3 letters at least. This way, the rates analyzers can recover data even from typing error. In the option used by the application, if the word search is temperature, the analysis could be made to tem, emp, mpe, per, era, rat, atu, tur and ura.

Another definition was chosen so that, regardless of typing the whole search was made in lowercase. Thus, if the word is "Temperature" the system switches to "temperature" and will do the analytical search

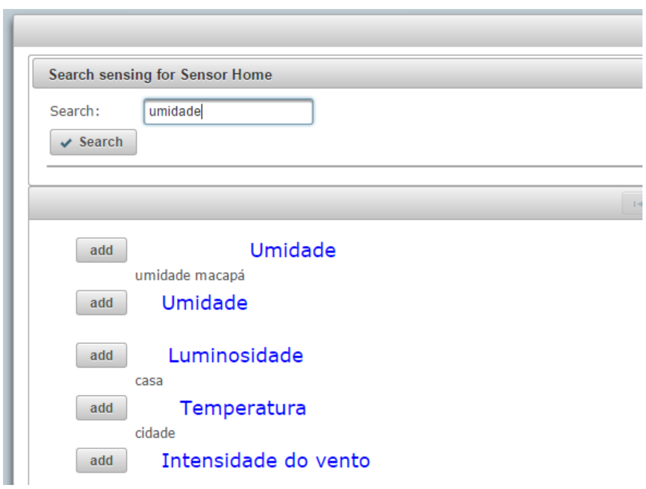

Figure 3. Sensor searcg screen

In Figure 3, it can check the implementation of the search functionality in use. In this example, the search word has "humidity". The search returns an objects list according to the relevance of what was searched. Thus, the first item in the list returned is a sensor named "humidity." As there were two registered, which has more fields of relevance is listed first. The first object contains both the word humidity in the name field, as in the annotation, while the second object only has this word in the name field.

The other results are listed by the analytical filter Ngram. The "Brightness" channel returned from the search "umi", "dad" and "ade", because this words combination is in the name field of this object, while the value "Temperature" was returned in the search because in its annotation field there is the word "city" that contains the Ngram, "dad" and "ade" atributes.

Therefore, with the library implementation, it is possible to perform the search of the measurement channels represented by the channel object in the database by analytical research into text.

D. Validation and Results

The final analysis and more important for class SaaS was sensors search. The parameter used to test was the response time and research relevance, as well as the data indexing time in the bank.

Three distinct databases were created for each scenario, to evaluate the sensor search performance, which is the main functionality for the end user, because it will serve as a data source for the sensors consumption as a service.

The scenarios were created as follows: scenario 1, with the amount of 9 sensors and 15 channels; scenario 2, with the amount of 99 908 sensors and 202 625 channels; and scenario 3, at the amount of 300,000 sensors and 3,000,000 sensor channels.

The search response time will serve directly to the end user because it will measure the time between the data request and the return of the time, which will influence the user experience relative to the solution.

The term "sensor recife" surveyed in Google returns a data set around 400,000 items, with the seek time of 650 milliseconds. This time was used as a reference to act on the user experience about extensively used service.

The time was classified into five levels, as follows:

- T1: simple search, with 1 word in 1 channel fields;
- T2: composed search, with 2 words in 1 channel fields;
- T3: composed search, with 2 words in 2 channel fields;
- T4: simple search, with 1 word with highest incidence in 1 channel fields;
- T5: composed search, with 2 words with highest incidence in 2 channel fields.

The search relevance analyses if the request returned to the user the data expected for him in the query.

The indexing time is a system parameter for assessing the time that the solution takes to index data in the database with the use of indexing through Hibernate Search. Table 3 shows the data obtained from the scenarios presented.

TABLE III. INDEXED QUERY SCENARIOS

Scenarios	T1	T2	T3	T4	T5
Scenario 1	8ms	15ms	16ms	11ms	16ms
Scenario 2	16ms	20ms	32ms	99ms	145ms
Scenario 3	21ms	28ms	41ms	139ms	283ms

The search time in three scenarios was well below the reference value for the proposal made for the solution. Thus, the user would receive your request, in the worst scenario, in 283ms.

 ISBN: 978-1-61208-438-1

The relevant factor is related to the quality of the data found with what the user was really looking for. Based on fuzzy feature, which delivers the term equal or similar to what was requested, it was asked to an users group a questionnaire that contained a description of the search task of sensing. Based on the results, the user would answer the questionnaire and then analyze it as shown in Table 4, with the following information, based on criteria listed below:

- Low: does not contain the requested data;
- Medium: contains the data similar to the requested;
- High: contains exactly the requested data.

TABLE IV. SEARCH RETURN RELEVANCE BY THE USER

Relevance level	P1	P2	P3
Low	0	0	0
Medium	0	1	0
High	7	6	7

Based on the data presented in the research referred to as P1, P2 and P3, the user found the requested data in the search.

The indexing time is a system parameter that influences indirectly in the sensors search. By means of this feature, the sensors are indexed and may be consumed by the search service more efficiently, through a text search. At the time, the solution is indexing only on its startup, however other strategies can be made so that this occurs at other times as well. The problem with indexing is that when it occurs, it consumes a lot of server performance, in addition to disable access to the base. One solution would be to perform indexing in minor peak times. However, as it is a cloud application and it can be accessed anywhere in the world, this time could not meet the solution, since the data flow in a particular location could be lower, but in another could be higher. The most recommended would be selective indexing, with the inclusion of new data as they were entering the base, but, over time, a full indexing would be required to keep the data as possible optimized on record.

TABLE V. SENSORS INDEX DATA

Scenarios	Total indexing time	Average indexed documents per second
Scenario 1	< 1ms	15
Scenario 2	57.025ms	3552,82
Scenario 3	280.234ms	3858,21

The indexing data were summarized in Table 5 with the following analysis:

- In scenario 1, the value was negligible, due to the low amount of data and, therefore, this value was expected
- Scenarios 2 and 3 showed indexing time values proportional to the size data in the database. Thus, the larger the database, the greater is the time for which such data to index.

These scenarios showed values close when considering the average of indexed documents per second. Such value is close because it is influenced by the server computing resources, as disk access time and processing. This value can be optimized by improving the cloud hardware.

IV. CONCLUSION

This paper presented a proposal for analysis in systems that work with sensors as a service, upon evaluation of the repository and data engine consumption. From this approach, it is possible to qualify systems that operate in this paradigm, setting a quality level and therefore provide the user with a better service. As future work other criteria can be added to assess the repository as characteristics related to performance, security, large amounts of data among others. This would increase the range of solutions that could be met from these analyses.

REFERENCES

[1] G. Press, "Internet of Things By The Numbers : Market Estimates And Forecasts", http://www.forbes.com/sites/gilpress/2014/08/22/internet-of-things-by-the-numbers-market-estimates-and-forecasts/, 2014 [retrieved: 03, 2015].

[2] H. Sundmaeker and A. Saint-exupéry, De, "Vision and Challenges for Realising the Internet of Things", CERPIoT, Luxembourg, 2010.

[3] ABI Research, "More Than 30 Billion Devices Will Wirelessly Connect to the Internet of Everything in 2020", https://www.abiresearch.com/press/more-than-30-billion-devices-will-wirelessly-conne, [retrieved: 11, 2014].

[4] H. Schaffers, N. Komninos, and M. Pallot, "Smart Cities as Innovation Ecosystems Sustained by the Future Internet", Fireball White Paper, EU, 2012.

[5] M. F. Catela, C. D. Pedron, and B. A. Macedo, "Service level agreement em cloud computing: um estudo de caso em uma empresa portuguesa". Univ. Gestão e TI. 4, 2014.

[6] N. W. Khang, "CLOUD COMPUTING SOLUTIONS: IAAS, PAAS, SAAS", http://wptidbits.com/techies/cloud-computing-solutions-iaas-paas-saas/, [retrieved: 08, 2014].

[7] G. Aceto, A. Botta, W de Donato, and A. Pescapè, "Cloud monitoring: a survey", Comput. Networks. 57, 2013, pp. 2093–2115.

[8] C. A. Kamienski, D. F. H. Sadok, E. M. Azevedo, R. A. M. B. K. Simões, and S. F. de L. Fernandes, "Um Modelo Integrado de Composição em Nuvem Computacional Conteúdo", Recife 2011.

[9] J. Simão and L. Veiga, "A classification of middleware to support virtual machines adaptability in IaaS", Proc. 11th Int. Work. Adapt. Reflective Middlew. - ARM '12, 2012, pp. 1–6.

[10] F. R. C. Sousa, L. O. Moreira, and J. C. Machado, "Cloud Computing: Concepts, Technologies, Applications and Challenges ", Escola Regional de Computação Ceará - Maranhão - Piauí., Parnaíba, 2009, pp. 25.

[11] A. Botta, W. Donato, V. Persico, and A. Pescap, "On the Integration of Cloud Computing and Internet of Things", FiCloud. Barcelona, 2014, pp. 8.

[12] A. A. Gavlak, L. Muratori, and D. A. Graça, " Image as a Service (IaaS): satellite images digital processing ZY-3 via web ". XXVI Congresso Brasileiro de Cartografia e V Congresso Brasileiro de Geoprocessamento. Gramado-RS 2014.

[13] X. Sheng, X. Xiao, J. Tang, and G. Xue, "Sensing as a service: A cloud computing system for mobile phone sensing". 2012 IEEE Sensors, 2012, pp. 1–4.

[14] P. Banerjee, C. Bash, R. Friedrich, P. Goldsack, B. A. Huberman, and J. Manley, "Everything as a service: Powering the new information economy", Computer (Long. Beach. Calif). 44, 2011, pp. 36–43.

[15] C. Perera, A. Zaslavsky, P. Christen, and D. Georgakopoulos, "Sensing as a Service Model for Smart Cities Supported by Internet of Things", Trans. Emerg. Telecommun. Technol. 2013, pp. 1–12.

[16] A. Manzoor, "Quality of Context in Pervasive Systems : Models, Techniques and Applications" Tu, Wien. 2010, pp. 155.

[17] J. G. Webster, "The measurement, instrumentation, and sensors handbook", CRC Press, 1999.

[18] JBoss, "Hibernate Search", http://docs.jboss.org/hibernate/search/3.4/reference/en-US/html_single/#preface, [retrieved: 01, 2015].