IOP Conf. Series: Materials Science and Engineering **144** (2016) 012005 doi:10.1088/1757-899X/144/1/012005

On performing semantic queries in small devices

C Costea, A Petrovan, L Neamț and O Chiver

Technical University of Cluj-Napoca, North University Centre of Baia Mare, Electrical Engineering, Electronics and Computers Department, 62A Dr. V. Babeș Street, 430083, Baia Mare, Romania

E-mail: ccostea@cunbm.utcluj.ro

Abstract. The sensors have a well-defined role in control or monitoring industrial processes; the data given by them can generate valuable information of the trend of the systems to which they belong, but to store a large volume of data and then analysis offline is not always practical. One solution is on-line analysis, preferably as close to the place where data have been generated (edge computing). An increasing amount of data generated by a growing number of devices connected to the Internet resulted in processing data sensors to the edge of the network, in a middle layer where smart entities should interoperate. Diversity of communication technologies outlined the idea of using intermediate devices such as gateways in sensor networks and for this reason the paper examines the functionality of a SPARQL endpoint in the Raspberry Pi device.

1. Introduction

The Internet of Things (IoT) refers to a global network of objects that are capable of sensing or acting, and able to communicate with each other, other machines or computers. IoT is defined by ITU and IERC as a “dynamic global network infrastructure with self-configuring capabilities based on standard and interoperable protocols where physical and virtual things have identities, physical attributes and virtual personalities, use intelligent interfaces and are seamlessly integrated into information network” [1]. The new term Industrie 4.0 [2] (or Industrial Internet [3]) refers to fourth industrial revolution; it is defined as “the integration of complex physical machinery and devices with networked sensors and software, used to predict, control and plan for better business and societal outcomes” [4]. In [2] are identified four key components: Cyber-Physical Systems, Internet of Things, Internet of Services, and Smart Factory.

A Cyber-Physical System (CPS) consists of a network of embedded devices (sensors, actuators) for monitoring and controlling of physical processes, including connectivity for both real-time data acquisition and information feedback to computational tools; these devices can capture large amount of data, supporting real-time control or analysis to reveals new insights and suggest actions. Programmable Logic Controllers (PLCs) have been used for decades in industrial process control [5], but in IoT trend PLC technologies are not competing with CPS, they complement each other: data from the PLCs can be transmitted to other entities (or data center) for statistical analysis to predict failures before they happen.

1.1. Edge Computing

Fog Computing was introduced by Cisco as a solution to a set of applications where real-time and latency-sensitivity is critical, extending the capabilities of Cloud resources to the edge of the network.

Fog Computing is a virtualized platform that provides [6] compute, storage, and networking services between end devices and traditional Cloud Computing Data Centers, typically, but not exclusively located at the edge of network. By example, in [7], data generated by wearables and other personal and medical devices is processed and stored in the Health-Fog node; the user can control the information, in particular determines with whom this data is shared.

The process of data sensor near the place at which it is sensed, converting it from a raw signal to contextually relevant information defines the edge mining (or data mining at the edge); by example [8] examines data mining on the wireless smart sensing devices that sit at the edge points of the Internet of Things.

Mobile Edge Computing (MEC) [9] aims to reduce both network latency and resource demands by shifting computing and storage capacities from the Internet cloud to the mobile edge. Services are hosted on MEC servers - operated by the mobile infrastructure provider. The proximity of the MEC servers to the mobile device increases responsiveness of mobile applications.

1.2. Internet of Things Gateways (IoTG)

In sensor networks, a gateway is a connection device between the (different) sub-networks with manager roles or for data concentration provided by a group of sensors. But sending data collected by sensors to a data center for further investigation is ineffective; therefore, it requires intermediate storage device to perform a preliminary processing (filtering, aggregation, annotation) before they are sent to the next hierarchical level.

An IoT Gateway is not only a proxy that transmits sensor data to a higher level - "IoT gateways are resource-constraint devices (i.e, with limited compute, memory, and storage amounts) which expose connected sensors as cloud services to become addressable, discoverable and controllable" [10].

1.3. *Ontologies in Sensor Networks*

The semantic sensor network use Semantic Web technologies and reasoning to interpret sensor data from physical devices that perform measurements. An example is the use of ontologies in clinical decision support systems (CDSS) [3]; by periodically collecting patient data and transferring them to physician is possible to supervision patient health status from remote sites. Smart Grid use sensors for advanced monitoring capabilities in real-time that can lead to more efficient management; by example phasor measurement units can sense grid instability within seconds [9] or smart meters installed at consumer locations for real-time bi-directional communication.

In [11] data distributed repositories are supported by distributed object caches with persistence capabilities between the components; it includes a semantic database storing event information to be directly processed by the agents or for semantic inferences.

2. Autonomy in Internet of Things

The concept of autonomy is determined by the ability of an agent to adapt its behavior; but do you really want a temperature sensor to change its behavior? However, the temperature can be supplied in different units, and this adaptation can be realized depending on the context, without an explicit user setting or transmitted parameters.

An important feature of device autonomy is related to context awareness. Compared to existing devices based on simple reflexes, an autonomous device should have the ability to choose their own actions without the intervention of another device or operator.

Due to the diversity of IoT technologies (consequently, a lot of proposed standards) and achieving the autonomy, semantic interoperability is emerging as an attractive solution for communication between entities (smart objects).

2.1. Semantic Interoperability in SSN

The IEEE defined in 1990 interoperability as "the ability of two or more systems to exchange data and to mutually understand the information which has been exchanged" [11].

The semantic sensor network use Semantic Web technologies and reasoning to interpret sensor data from physical devices that perform measurements.Sensor data can be annotated with semantic metadata to provide contextual information, making sensor descriptions and measurements available on the Web. Publishing them on LOD (Linked Open Data) cloud enables finding other related data and facilitates integration of data from different sources [12].

Sheth et.al. [6] annotated sensor data with semantic metadata and proposed the Semantic Sensor Web (SSW). In [13] are identified three most important directions on SSW: automatically annotate sensor data, by providing semantic metadata; publish annotated sensor data using vocabularies and schemas to enable discovery; reasoning on semantically enriched sensor data.

Through semantic annotation metadata are associated to measured values (i.e., sensor capabilities, time, location, etc.). These tags represent concepts, properties and relationships and should conform to ontology.

The W3C Semantic Sensor Network Incubator Group developed an ontology [14], [15] to describe sensors and sensor networks for use in sensor network and sensor web applications and recommended methods for developing applications according to the Open Geospatial Consortium's (OGC™) Sensor Web Enablement (SWE) standards.

Linked Sensor Data (LSD) is achieved by transforming raw sensor observations to RDF format and by linking it with other datasets on the Linked Open Data (LOD) cloud.

In [11] data distributed repositories are supported by distributed object caches with persistence capabilities between the components; it includes a semantic database storing event information to be directly processed by the agents or for semantic inference.

Resource Description Framework (RDF) is a directed, labeled graph data format for representing information in the Web. SPARQL [16] is a semantic query language able to retrieve and manipulate data stored in Resource Description Framework (RDF) format; its protocol allows a query to be sent to a remote service endpoint and the results sent back (in RDF, XML, JSON). SPARQL can be used to express queries across diverse data sources, whether the data is stored natively as RDF or viewed as RDF via middleware. Federated Query [17] is an extension of SPARQL 1.1 (and W3C Recommendation) for executing queries distributed over different SPARQL endpoints.

2.2. Linked Stream Middleware

Linked Stream Data extended the RDF data model for representing stream data generated from sensors or social network applications. The identification of events that can potentially affect the normal functioning can be achieved by query processing over semantic streams; a detected event triggers and determines which adaptation is needed.

Several platforms have been proposed as processing engines for LSD, including Streaming SPARQL, C-SPARQL or CQELS. The whole embedded CQELS is smaller than 10MB and needs only 4-6MB of RAM to process millions of triples on various small devices [18].

2.3. Interoperability by Ontologies

An example is the use of ontology in Clinical Decision Support Systems [19]; by periodically collecting patient data and transferring them to physician is possible to supervision patient health status from remote sites.

Smart Grid use sensors for advanced monitoring capabilities in real-time that can lead to more efficient management; by example [10] phasor measurement units can sense grid instability within seconds or smart meters installed at consumer locations for real-time bi-directional communication.

Stream Annotation Ontology (SAO) allows representation of aggregated stream data and temporal characteristics.

3. Experiment

To check the possibilities of implementing a storage capable intermediate and preliminary filtering of the data we used a sensor temperature (wireless XRF Ciseco sensor) and a Raspberry Pi system

IOP Conf. Series: Materials Science and Engineering **144** (2016) 012005 doi:10.1088/1757-899X/144/1/012005

(ARM1176JZF-S 700 MHz processor, 512MB memory), Raspbian (Linux) operating system, Oracle Java7u40 and Jena Fuseki server. To evaluate the performance of smart sensor network for distributed analyzing a simple architecture was created which includes Ciseco temperature sensors, Raspberry Pi (RPi) device connected to laboratory local network, laptop and Android tablet PC (Samsung Galaxy Tab E).

The Raspberry Pi is a low cost, credit-card sized computer that plugs into a computer monitor or TV, and uses a standard keyboard and mouse. RPi has the ability to interact with the outside world, and has been used in a wide array of digital maker projects, from music machines to weather stations.

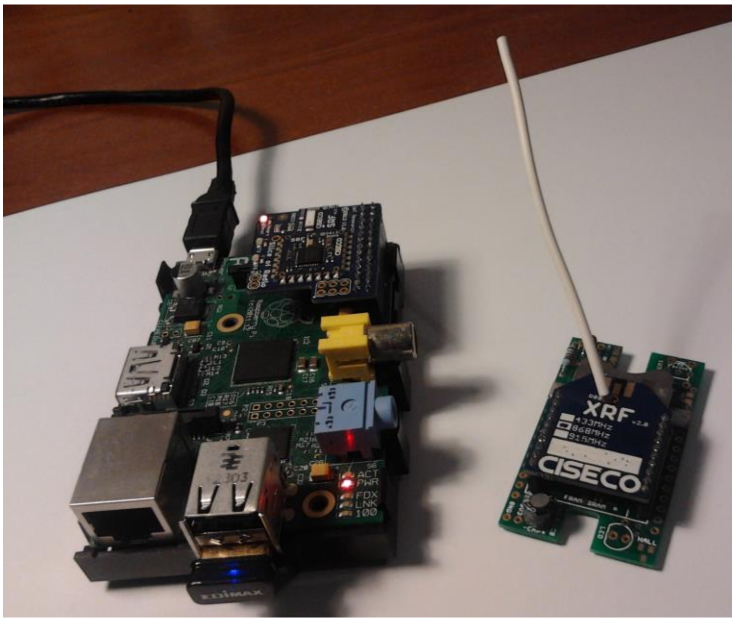

Figure 1. Experiment wih RPi and Ciseco temperature sensor

Linked Sensor Data (LSD) is achieved by transforming raw sensor observations to RDF format and by linking it with other datasets on the Linked Open Data (LOD) cloud. Linked Sensor Data (LSD) is achieved by transforming raw sensor observations to RDF format and by linking it with other datasets on the Linked Open Data (LOD) cloud.

We used a combination of Slice of Pi, a temperature XRF sensor and XRF Wireless RF Radios from Ciseco who promotes Lightweight Local Automation Protocol (LLAP).

The remote sensors XRF must be configured to capture the temperature from the onboard thermistor and transmit it; device sleep most of the time to conserve energy and periodically wake up and transmit readings; the XRF also has the ability to send battery levels. Ciseco have developed firmware that can be loaded onto the RF modules to give them different "personalities" (temperature sensor, hall-effect, button, relay) - specific message sets for LLAP devices.

The experiments used an Android tablet PC for graphical representations of time series dataset gathered from gateway accessed through REST services, Figure 2.

Jena use various storage mechanisms: in-memory, relational database (SDB) or disk based tuple index (TDB). SDB uses an SQL database for the storage and query of RDF data. TDB can be used for RDF storage and query as a high performance RDF store on a single machine. Jena stores information as RDF triples in directed graphs. Fuseki is an add-on that provides a SPARQL endpoint via an HTTP interface. Fuseki acts as a data publishing server, thus allowing RDF graphs to be accessed and updated using SPARQL and HTTP.

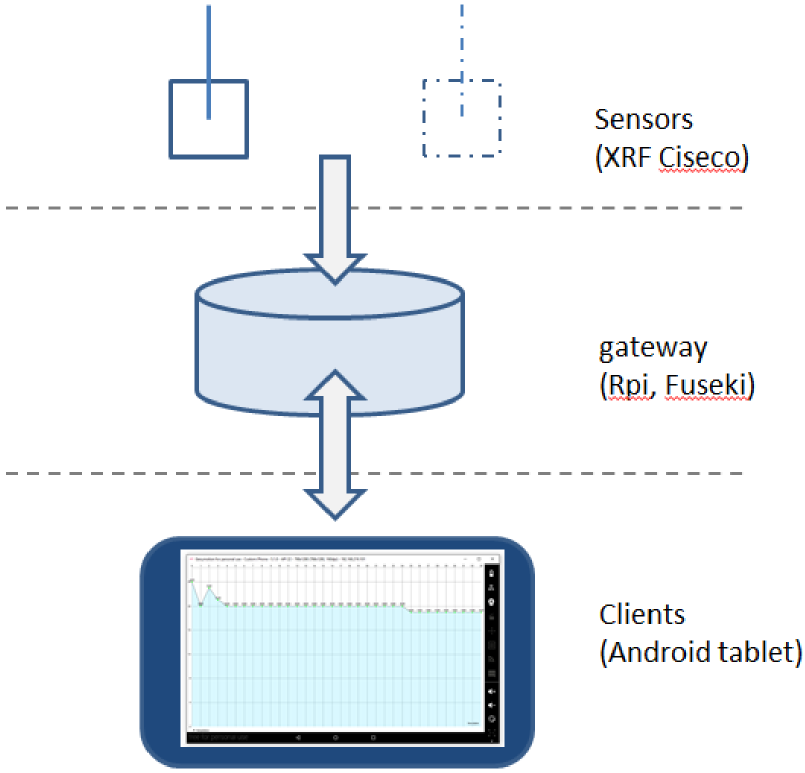

Figure 2. Data workflow of the experiment

In our preliminary experiments we are interested to evaluate in-memory mechanism in small devices; we consider that the device layer of IoT applications includes nodes for temporary data storage, filtering and annotating data sensors (such as gateway functions in WSN).

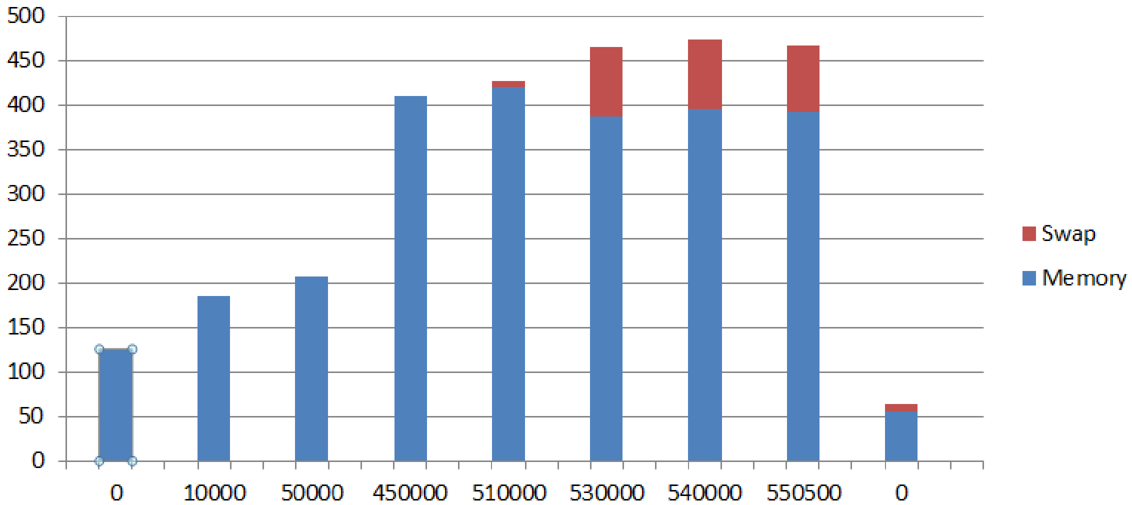

Figure 3. Number of triples and memory used in RPi device

We inserted triplets in the RDF graph until the full memory device, but maintaining its functionality; the final dataset contained over 550,000 nodes; however a simple SPARQL query for counting the nodes lasted 27 seconds.

Figure 3 and Table 1 show memory usage; swap memory started by 500,000 nodes (free internal memory was 17 M). In Table 1 see the increasing response time for just two SSH connections on device and one client accessing Fuseki server. During the test was possible to run Ruby command line scripts provided by Fuseki server.

Table 1. Results - response time and used memory

Dataset size (nodes)	Query response (msec)	Memory (MB)
1000	392	180
2000	506	184
50000	2453	208
500000	26359	474

```
<http://cti.ubm.ro/sensors/datastream/Temp#observations_point_34>
        a ss:Point ;
        ss:hasUnitOfMeasurement  unit:degree-Celsius ;
        ss:time [ a       tl:Instant ;
                    tl:at   "2016-01-29T18:55:47"^^xsd:dateTime
                  ] ;
        ss:time [ a       tl:Instant ;
                    tl:at   "2016-01-29T19:54:36"^^xsd:dateTime
                  ] ;
        ss:value  "21.18" , "20.91" ;
        ns:featureOfInterest
<http://cti.ubm.ro/sensors/datastream/Temp#context_one> .

<http://cti.ubm.ro/sensors/datastream/Temp#context_one>
        a  ss:FeatureOfInterest ;
        ct:hasFirstNode  [ a                 ct:Node ;
                           ct:hasLatit   47.660331 ;
                           ct:hasLongit  23.545122 ;
                           ct:hasNodeName   "NUCBM"
                         ] .
```

Figure 4. A sample measurement data represented in Fuseki server

4. Conclusion

It is not comparable this experiment with existing benchmarks results for SPARQL engines; however, these preliminary results allow a distributed approach (based on REST services), in which multiple software entities can share knowledge, according to LOD principles.

The paper explore the decentralize paradigm based on data mining technologies and smart devices processing data locally instead of centralized to a data center. We investigate applications of distributed analysis in smart grid context and present preliminary experiments of sensor data acquisition with RPi devices, cloud storage through REST services and a mobile Android application.

This functional model will be further used in various simulations and to study specific frameworks or distributed algorithms. The experimental results revealed the usefulness of interim storage and online processing/analysis (or even real-time) of data streams simultaneously with the filtered data storage in the cloud. With the significant increased capabilities of embedded devices, agents and mobile applications more flexible smart sensors can be achieved and distributed data analysis can be part of emerging grid technologies.

References

[1] Atzori L, Iera A and Morabito G 2010 The Internet of Things: A survey, *Computer Networks* **54**(15) 2787-2805

[2] Sheth A, Henson C and Sahoo S S 2008 Semantic Sensor Web, *IEEE Internet Computing* **12**(4) 78-83

[3] Barnaghi P and Presser M 2010 *Publishing linked sensor data*, 3rd International Workshop on Semantic Sensor Networks CEUR-WS 2010, Shanghai, China, November 7, pp 1-16

[4] Gaura E I, Brusey J, Allen M, Wilkins R, Goldsmith D and Rednic R 2013 Edge Mining the Internet of Things, *IEEE Sensors Journal* **13**(10) 3816-3825

[5] Barz C, Deaconu S I, Latinovic T, Berdie A, Pop-Vadean A and Horgos M 2015 PLCs used in smart home control, *IOP Conf. Ser.: Mater. Sci. Eng.* **106** 012036

[6] Bonomi F, Milito R, Zhu J and Addepalli S 2012 *Fog Computing and Its Role in the Internet of Things,* First edition of the MCC workshop on Mobile cloud computing MCC '12, Helsinki, Finland, August 13-17, pp 13-16

[7] Ferrer-Roca O, Roca D, Nemirovsky M and Milito R 2015 *HFOG: Small versus Big Data*, The International eHealth, Telemedicine and Health ICT Forum for Educational, Networking and Business, Luxembourg, April 22-24, pp 409-415

[8] Muhammad Habib ur Rehman, Chee Sun Liew, Teh Ying Wah, Junaid Shuja and Babak Daghighi 2015 Mining Personal Data Using Smartphones and Wearable Devices: A Survey, *Sensors* **15**(2) 4430-4469

[9] Kolozali S, Bermudez-Edo M, Puschmann D, Ganz F and Barnaghi P 2014 *A Knowledge-Based Approach for Real-Time IoT Data Stream Annotation and Processing*, IEEE International Conference on Internet of Things (iThings), and IEEE Green Computing and Communications (GreenCom) and IEE Cyber, Physical and Social Computing (CPSCom), Taipei, Taiwan, September 1-3, pp 215-222

[10] Zhou Q, Natarajan S, Simmhan Y and Prasanna V 2012 *Semantic Information Modeling for Emerging Applications in Smart Grid*, 9th International Conference on Information Technology: New Generations (ITNG), Las Vegas, USA, April 16-18, pp 775-782

[11] Penya Y K, Nieves J C, Espinoza A, Borges C E, Pena A and Ortega M 2012 Distributed Semantic Architecture for Smart Grids, *Energies* **5**(11) 4824-4843

[12] Azhar Mahmood, Ke Shi, Shaheen Khatoon and Mi Xiao 2013 Data Mining Techniques for Wireless Sensor Networks: A Survey, *International Journal of Distributed Sensor Networks* **2013** 406316

[13] Vermesan O and Friess P 2014 *Internet of Things - From Research and Innovation to Market Deployment*, River Publishers, Aalborg, Denmark

[14] Moraru A and Mladenic D 2012 *A framework for semantic enrichment of sensor data*, 34th International Conference on Information Technology Interfaces (ITI 2012), Dubrovnik, Croatia, June 25-28, pp 155-160

[15] Jesus Arias Fisteus, Norberto Fernandez Garcia, Luis Sanchez Fernandez, Damaris Fuentes-Lorenzo 2014 Ztreamy: a middleware for publishing semantic streams on the Web, *Journal of Web Semantics* **25**(1) 16-23

[16] Le-Phuoc D, Dao-Tran M, Le Van C, Le Tuan A, Nguyen Duc M, Tran Nhat T and Hauswirth M 2015 *Platform-Agnostic Execution Framework - Towards RDF Stream Processing*, RDF Stream Processing Workshop at ESWC2015, Portoroz, Slovenia, May 31, P10

[17] Compton M, Barnaghi P, Bermudez L, Garcia-Castro R, Corcho O, Cox S, et.al 2012 The SSN ontology of the W3C semantic sensor network incubator group, *Web Semantics: Science, Services and Agents on the World Wide Web* **17**(1) 25-32

[18] Abbasi A A and Younis M 2007, A survey on clustering algorithms for wireless sensor networks, *Computer Communications* **30**(14-15) 2826-2841

[19] W3C Incubator Group Report, *Semantic Sensor Network XG Final Report*-2011 [online], http://www.w3.org/2005/Incubator/ssn/XGR-ssn-20110628/