

# Assessing RDF Graph Databases for Smart City Services

Pierfrancesco Bellini, Paolo Nesi

Distributed Systems and Internet Technology Lab, DISIT, <http://www.disit.org>  
Department of Information Engineering, DINFO, University of Florence, Florence, Italy  
[pierfrancesco.bellini@unifi.it](mailto:pierfrancesco.bellini@unifi.it), [paolo.nesi@unifi.it](mailto:paolo.nesi@unifi.it)

**Abstract** — RDF stores may be used to set up knowledge bases integrating heterogeneous information for web and mobile applications to use the data for new advanced services to citizens and city administrators, thus exploiting inferential capabilities, temporal and spatial reasoning, and text indexing. In this paper, the needs and constraints for RDF stores to be used for smart cities services, together with the currently available RDF stores are evaluated. The assessment model allows a full understanding of whether they are suitable as a basis for Smart City modeling and application. The comparison of the RDF stores addressed a number of well-known RDF stores. The paper also reports the adoption of the proposed *Smart City RDF Benchmark* on the basis of Florence Smart City model, data sets and tools accessible as Km4City <http://www.km4city.org>, and adopted in the European Commission international smart city projects named RESOLUTE H2020, REPLICATE H2020, and in Sii-Mobility National Smart City project in Italy.

**Keywords**— *smart city; RDF stores; graph databases; RDF benchmark; linked data benchmark.*

## I. INTRODUCTION

Smart cities produce large amount of data having a large variability, variety, velocity, and size; and thus complexity. The variety and variability of data can be due to the presence of several different formats, [1], [2] and to the interoperability among semantics of the individual fields and of the several data sets [3]. Static data are rarely updated, for instance once per month/year, which is quite the opposite with dynamic data: they can be updated from once a day up to every minute so as to get real time data. The data velocity is related to the frequency of data update for dynamic data such as position of buses, people flow status, position of waste collectors, etc. The size grows over time accumulating new data every day and week. At architectural level, smart city solutions typically adopt n-tier architectures [4].

The usage of RDF stores in the application domain of Smart City is quite recent, since in most cases services are vertically provided. For example the Intelligent Transport System, ITS, in the city provides information regarding the location of buses and their delay, without addressing the location of city services, flow of people, real time events in the city. Some city data integrators are well-known services such as bike and car sharing, navigator system, tourism information, hotel booking, etc. All these solutions have the need to integrate geo-located information with real time data and events continuously arriving from updated information such as: events, votes, traffic flows, comments, etc. [2], [5]. As to these

applications, RDF stores may be a solution to allow addressing the variability of data, so as to make reasoning on space, time, and concepts [6]. The Resource Description Framework specified by W3C allows the representation of facts using “triples” of the form (*subject, predicate, object*) where URIs are used to identify the entities and the predicates connecting them. Thus a triple represents the arc of the graph connecting two entities and the predicate describes the kind of relation between the two entities. Moreover the object part of the triple can also be a low level data type as string, dates, integers etc. to describe not only the relations among entities but also specific information about them (e.g. name, email, birth date). RDF stores allow storing these triples and the SPARQL query language allows querying them. Some RDF stores can also manage set of triples as a single graph identified by an URI, in this way information about this graph can be provided using other triples (where the subject is the graph itself).

For the evaluation of RDF stores specific assessment models and benchmarks have to be adopted. For example, the LUBM benchmark [7] uses a synthetic dataset in the university domain and covers only the SPARQL 1.0 specification. On the contrary, the BSBM benchmark [8] generates a synthetic dataset in the e-commerce domain and covers the SPARQL 1.1 business analytics queries. More recently, in the Linked Data Benchmarks Council project<sup>1</sup> two benchmarks were proposed both generating a synthetic dataset, one from the semantic publishing domain (LDBC-SP) and the other from the social networks domain (LDBC-SN). The GeoSPARQL standard has been developed by the Open Geospatial Consortium to cover spatial searches, while not many solutions currently support this specification. Regarding the benchmark of geo and spatial RDF stores the Geographica benchmark [9] was proposed by using both a synthetic generated dataset and a real dataset. It analyses the support and performance for advanced spatial relationships among complex spatial entities (e.g., polygons). In [10], the real and synthetic benchmark datasets have been compared showing that synthetic generated datasets are similar to datasets generated for relational database benchmarks (TPC-H) and strongly different from real-world datasets (e.g., dbPedia) being much less structured. In [11], with the SPARQL Performance Benchmark (SP2Bench) a language-specific benchmark framework designed for the most common SPARQL constructs has been proposed.

Recently SPARQL has been extended to query real-time data coming from RDF data streams. There are some

---

<sup>1</sup> <http://ldbncouncil.org>

implementations as C-SPARQL [12], SparqlStream [13], CQELS and also specific benchmarks were defined as SRBench [14] using data from weather sensors, LSBench [15] using data from social networks and CityBench [16] using data from smart city sensors. Those kinds of specific benchmark are suitable for streaming data, with queries performing specific requests with limited number of results. W3C also reviewed RDF store benchmarks<sup>2</sup> highlighting their applicability in assessing different aspects of the RDF stores, and their application on different stores.

Despite this wide state of the art on RDF stores benchmarks, none of the mentioned approaches is specifically suitable for assessing the RDF stores against Smart City. Smart City presents extremely particular and specific conditions exploiting the latest and most challenging constructs of the RDF stores as geo-spatial queries, text queries, time queries and combinations of them. On this regard, in this paper, a Smart City RDF Benchmark, SCIRB, has been.

This paper reports the formalization of the proposed *Smart City RDF Assessment Model and Benchmark* and its adoption in comparative assessment of a number of RDF stores. The data and queries adopted for replicating the mentioned assessment have been published on the following web page <http://www.disit.org/smartcityrdfbenchmark>. The dataset is real and is based on Florence Smart City which in turn is grounded on Km4City ontology and model [3].

The paper is structured as follows. In **Section II**, the major smart city requirements/demands in modeling and accessing semantic knowledge are reported. The requirements can be used as drivers for features based selection of RDF stores. **Section III** presents the general evaluation methodology for assessing and selecting the RDF stores for smart city applications. In **Section IV**, the comparison of most relevant state of the art RDF stores is reported on the basis of the model identified in Section III.A. **Section V** reports the application of the proposed smart city benchmark in assessing the most featured RDF stores (i.e., Virtuoso, GraphDB, Oracle and StarDog). The analysis has highlighted several interesting aspects connected to the performance of RDF stores in: loading and indexing triples, and in performing geographical and textual queries, also during store updates. Conclusions are drawn in **Section VI**.

## II. SMART CITY REQUIREMENTS FOR RDF STORES

When providing services to citizens of a smart city, an RDF/graph store should provide some features that allow the support of specific functionalities. In particular, the following features are reported according to their relevance and classifying them. Therefore, smart city stores should provide support for:

- *spatial indexing (must have)*: providing information near to a given geographical point: as a GPS location. For example, all the services that are currently closed/unavailable to a given point. It should also support advanced geo-spatial functionalities as being able to

manage complex geometries (e.g., information along a cycle path, all elements into a given polygonal area).

- *high performance on spatial querying*.
- *full text indexing (must have)*: allowing the integration of semantic queries with keyword based searches on text which can be present into the attributes and class elements, as triples. Subjects and objects of triples can contains relevant text area such as descriptions, street names, locations names, etc.
  - *high performance on full text querying*.
- *quadruples* (not only triples) to associate dataset metadata with the loaded triples (*must have*). Triples are produced on the basis of data coming from many different sources. Therefore, it is important to track the data source, with metadata and associated licenses. This feature is particularly useful to solve or process licenses during the data usage from clients and via APIs.
- *some kinds of inference* (good to have) such as the basic RDFS or the more advanced OWL2 profiles allowing the inference of new facts from the available data. This may be used to generalize/specialize about entities, to same-as, equivalence, transitive, symmetrical, etc. The inference may imply the materialization of triples in the phase of indexing [Bellini et al., 2015].
- *temporal indexing* (good to have): many information and features are changing over time in smart cities. For example, the weather situation and its related forecast, the traffic flow detected from traffic sensors, the position of buses, and events occurring within the city. For this reason, it is quite important that the RDF store should support temporal search to allow the easy retrieval of temporal data. Moreover, the storing of temporal data (that may change in real time) is the main source for increasing the database size, demanding big data solutions for smart city for volume, velocity and variety, at least.
- *high volume of queries* (good to have). Dealing with bigdata RDF store with many users querying the data is quite challenging, for this reason a clustering solution is needed. It could be a clustering (vertical scale or scale up/down) when the same service is duplicated to allow many concurrent queries and to provide also a fault tolerance solution. It could be also a scale out clustering (horizontal) when data are split among different servers, as a single server cannot handle all the data.

A very relevant non-functional requirement is due to the fact that when it comes to Smart City applications, they are often exploited by Public Administrations. They ask for: (i) standard solution to avoid the risk of vendor lock-in especially for very new technologies like RDF stores are; (ii) open source solutions to be compliant with typical national laws encouraging open solutions with source code accessible and shareable among several public administrations. Moreover, there should be an active community handling and supporting the product.

<sup>2</sup> <https://www.w3.org/wiki/RdfStoreBenchmarking>

### III. EVALUATION METHODOLOGY

The *Smart City RDF Assessment Model and Benchmark* evaluation methodology is carried out within two phases.

In the **first phase** the *Smart City RDF Assessment Model* is applied. It consists in an analysis of some general features according to the requirements provided in Section II, and more particularly to verify if the RDF/graph store provides support for: SPARQL v.1.1, inference, triples or quadruples, etc.

In the **second phase**, the *Smart City RDF Benchmark* is applied. It is based on performance tests grounded on a set of SPARQL queries designed by considering all the aspects, and including spatial and full text searches (in many cases the SPARQL queries have been designed by adopting the specific constructs related to the different stores). The execution of the Benchmark consists in assessing the performance on the identified queries on three datasets with growing size expanding temporal horizon (1 month, 2 months and 3 months of cumulated real-time data).

#### A. Smart City RDF Assessment Model

As to the Smart City RDF assessment model, the features considered to analyse the RDF stores are the following:

- *SPARQL version* supported being 1.0 or 1.1;
- inference type supported as full materialization of triples at load time or materialization at query time, and the inference profiles supported (e.g., RDFS, RDFS+, OWL, OWL2, OWL2-DL, ...);
- If the store is a triple or quadruple store, check whether it stores only the *subject predicate object* or it can have also a *context* URI;
- How the triples/quadruples are physically stored, namely by using a custom indexing or an RDBMS or other external service (e.g., HBase, Cassandra);
- If the store supports *Clustering* where replicated nodes are used for high availability and fault tolerant solution;
- If the store supports *Scale Out Clustering* where data are allocated on multiple nodes, while no node contains all the data (index sharing);
- If the store supports *Spatial search* at Basic level (meaning that it is able to index and retrieve only geolocated points) or at Advanced level (meaning that it is able to index complex shapes, for example polylines);
- If the store supports full text search, providing the ability to search using keywords;
- If the store allows the association of triple/quadruples with a temporal validity contexts, thus allowing to easily filter triples by means of temporal constraints;
- Size of stores managed as the largest number of triples/quadruples reported to be managed by the RDF store in the literature;
- License under which the RDF store is available, being either open source or commercial;

TABLE I. QUERIES OF SMART CITY RDF BENCHMARK: QUERY NAME, DESCRIPTION AND IF THE QUERY EXPLOIT INFERENCE OR NOT.

Query	Description	infer- ence	Geo- spat.	Full- text
<i>Find-address</i>	given the latitude and longitude position it retrieves the nearest address within 100m.	No	Yes	No
<i>Municipalities-florence</i>	It retrieves the list of municipalities within the province of Florence.	No	No	No
<i>Bus-lines</i>	It retrieves the list of bus lines.	No	No	No
<i>Bus-stops-of-line</i>	given the bus line, it retrieves the complete bus stop list of the line.	No	No	No
<i>Lines-of-bus-stop</i>	given a bus stop, it retrieves the lines going past that bus stop.	No	No	No
<i>Bus-stop-latLng</i>	given a position and a radius, it finds the bus stops that are within the radius.	No	Yes	No
<i>Bus-stop-florence</i>	It retrieves all the bus stops in the municipality of Florence.	No	No	No
<i>Bus-stop-forecast</i>	given a bus stop, it finds the next forecasts for the lines going past that bus stop.	No	No	No
<i>AVM-distribution</i>	It retrieves for each day the count of the received AVM records.	No	No	No
<i>Service-florence</i>	It retrieves all the services in the municipality of Florence.	Yes	No	No
<i>Service-Acc-Clt-Trs-W&amp;F-florence</i>	It retrieves all the services in the Accommodation, Cultural Activity, TourismService and Wine&Food classes within the municipality of Florence.	Yes	No	No
<i>Service-Htl-B&amp;B-florence</i>	It retrieves all the services in the Hotel and Bed&Breakfast classes within the municipality of Florence.	Yes	No	No
<i>Service-latLng</i>	It retrieves the services within a radius from a latitude, longitude position.	Yes	Yes	No
<i>Service-Acc-Clt-Trs-W&amp;F-latLng</i>	It retrieves all the services in the Accommodation, Cultural Activity, TourismService and Wine&Food classes within a radius from a position.	Yes	Yes	No
<i>Service-Htl-B&amp;B-latLng</i>	It retrieves all the services in the Hotel and Bed&Breakfast classes within a radius from a given position.	Yes	Yes	No
<i>Full-text</i>	It retrieves anything matching a keyword	No	No	Yes
<i>Service-text-florence</i>	It retrieves all the services in the municipality of Florence matching a keyword.	Yes	No	Yes
<i>Service-text-latLng</i>	It retrieves all the services matching a keyword given a position and a radius.	Yes	Yes	Yes
<i>Sensor-florence</i>	It retrieves all the sensors within the municipality of Florence.	No	No	No
<i>Sensor-latLng</i>	It retrieves all the sensors within a radius from a position.	No	Yes	No
<i>Sensor-status</i>	It retrieves the latest information associated with a sensor.	No	No	No
<i>Sensor-distribution</i>	It finds for each day the count of the received sensor status updates.	No	No	No
<i>Parking-status</i>	It retrieves the latest information associated with a parking lot.	No	No	No
<i>Parking-distribution</i>	It retrieves for each day the count of the acquired parking status records.	No	No	No
<i>Weather-florence</i>	It retrieves the latest forecast available for the municipality of Florence.	No	No	No
<i>Weather-distribution</i>	It retrieves for each day the count of the acquired weather forecasts.	No	No	No

- Development language (e.g., Java, C);
- If the project is still an active project, date of last activity, date of last release;

Detailed performance testing should be performed on stores that support minimum set of requirements and in particular providing at least support for:

- SPARQL 1.1 as it provides aggregation functions (group by, count) and other features that were missing in 1.0;
- RDFS inference at load time or query time;
- Quadruples, so that correct metadata can be associated with datasets;
- basic spatial search to allow searching services via position;
- full text search to be able to integrate keyword search with semantic search;
- “Big stores” management in some how: that is the capability of managing large data store with some technique, scaling for instance.

If the RDF store supports additional features, they are positively considered in the context.

### B. Smart City RDF Benchmark

In this section, the queries at the basis of the *Smart City RDF Benchmark* are presented. The queries performed over the datasets are mainly the ones behind a real Smart City application and the API adopted in Km4City and used in <http://servicemap.km4city.org>. ServiceMap is an accessible smart city web application for developers to develop informative totems, while the Km4City API is a set of services accessible from Smart City mobile app delivered on all the available platforms: Apple Store, Google Play, and Windows Market.

Noteworthy is that the SPARQL recommendation does not cover the geo-spatial queries, nor the full-text queries. Therefore, in order to support those features, RDF store builders/vendors implemented these features by using their own specific syntax. Due to this reason, for some queries there is not a unique formulation and the query has to be adapted for each RDF store under test (they can be accessed from the web page of the proposed benchmark <http://www.disit.org/smartcityrdfbenchmark>). In Table I, the semantic queries at the basis of the *Smart City RDF Benchmark* are described and what is highlighted is whether the single query involves in its definition according to the ontology the exploitation of: inference<sup>3</sup>, geo-spatial and/or full-text aspects.

For example, the query to retrieve the last weather forecast for the municipality of Florence is the following:

```
PREFIX ...
SELECT ?day ?desc ?minTemp ?maxTemp ?time ?wPred
WHERE {
  {
    SELECT DISTINCT ?wRep ?time WHERE {
      ?munic rdf:type km4c:Municipality;
      foaf:name "FIRENZE";
      km4c:hasWeatherReport ?wRep.
      ?wRep km4c:updateTime/schema:value ?time.
    } ORDER BY DESC(?time) LIMIT 1
  }
  ?wRep km4c:hasPrediction ?wPred.
  ?wPred dct:terms:description ?desc;
  km4c:day ?day;
  km4c:hour "giorno"^^xsd:string.
```

```
OPTIONAL { ?wPred km4c:minTemp ?minTemp.}
OPTIONAL { ?wPred km4c:maxTemp ?maxTemp.}
}
```

It uses a sub-query to find the last report received related to the municipality and from this the prediction associated is selected and the associated information is returned.

A query using full text search and geospatial proximity search (using the syntax of virtuoso) is:

```
PREFIX ...
SELECT DISTINCT ?ser ?elong ?elat ?sTypeIta WHERE {
  ?ser ?p ?txt.
  ?txt bif:contains "casa".
  {
    ?ser km4c:hasAccess ?entry.
    ?entry geo:lat ?elat;
    geo:long ?elong;
    geo:geometry ?geo.
    filter(bif:st_intersects(?geo,
      bif:st_point(11.26193046,43.77072194), 0.5))
  } UNION{
    ?ser geo:lat ?elat;
    geo:long ?elong;
    geo:geometry ?geo.
    filter(bif:st_intersects(?geo,
      bif:st_point(11.26193046,43.77072194), 0.5))
  }
  ?ser a ?sType.
  FILTER(?sType!=km4c:RegularService &&
    ?sType!=km4c:Service)
  ?sType rdfs:label ?sTypeIta.
  FILTER(LANG(?sTypeIta)="it")
}
```

As it occurs with all the RDF benchmarks, the SPARQL queries are specifically tuned for a model. In this case, queries have been designed for the model described in the next section. The complete formalization of the queries, as well as the dataset dumps adopted in the tests reported hereafter, are available at

<http://www.disit.org/smartcityrdfbenchmark>

TABLE II. DATASETS CHARACTERIZATION FOR SMART CITY BENCHMARK.

Type	1 month		2 months		3 months	
	<i>Triples</i>	%	<i>triples</i>	%	<i>triples</i>	%
AVM	8.4M	19%	18M	33%	28M	43.1%
Parking	413k	0.9%	976k	1.8%	1.4M	2.1%
Sensors	900k	2%	1.7M	3.1%	2.2M	3.3%
Weather forecast	15k	0%	23k	0%	23k	0%
<b>Total dynamic</b>	<b>9.7M</b>	<b>22%</b>	<b>21M</b>	<b>38%</b>	<b>32.5M</b>	<b>48.5%</b>
Road graph	33.5M	75%	33.5M	60.3%	33.5M	50%
Services	681k	1.5%	681k	1.2%	681k	1%
Other static	286k	0.6%	286k	0.5%	286k	0.4%
<b>Total static</b>	<b>34.5M</b>	<b>78%</b>	<b>34.5M</b>	<b>62%</b>	<b>34.5M</b>	<b>51.4%</b>
<b>Total</b>	<b>44.2M</b>	<b>100%</b>	<b>55.6M</b>	<b>100%</b>	<b>67.5M</b>	<b>100%</b>

<sup>3</sup> <https://www.w3.org/TandS/QL/QL98/pp/queryservice.html>

TABLE III. RDF STORES' FEATURES COMPARISON  
where: OS=Open Source, Cm=Commercial, H=Horizontal cluster, V=Vertical cluster

RDF Store	Features											
	SPARQLv.	Inference	3/4-uple	Spatial search	Full-text search	Storage	Temporal search	Size (million/billion triples)	License	Dev. Language	Cluster support	Active project
<b>Virtuoso 7.2.4 OS</b>	1.1	RDFS+	4	Adv	Y	RDBMS	N	50BT	OS	C	N	Y
<b>Virtuoso 7.2.4 Comm</b>	1.1	RDFS+	4	Adv	Y	RDBMS	N	50BT	Cm	C	H	Y
<b>Graph DB SE 7.0.1</b>	1.1	OWL2RL	4	Bas	Y	custom	N	10BT	Cm	Java	H	Y
<b>Stardog 4</b>	1.1	OWL2	4	Adv	Y	custom	N	10BT	Cm	Java	H	Y
<b>Oracle 12c</b>	1.1	RDFS, OWL2	4	Adv	Y	custom	N	1TT	Cm	C/Java	H	Y
Apache Jena-Fuseki	1.1	RDFS OWL-Lite	3	Bas	Y	custom (TDB)	N	1.7BT	OS	Java	N	Y
Apache Jena-Fuseki	1.1	No	4	Bas	Y	custom (TDB)	N	1.7BT	OS	Java	N	Y
Blazegraph 2.1.2	1.1	RDFS+	3	Bas	Y	custom	N	50BT	OS	Java	V&H	Y
Blazegraph 2.1.2	1.1	No	4	Bas	Y	custom	N	50BT	OS	Java	V&H	Y
CumulusRDF	1.1	No	3	No	N	Cassandra 1.2	N	120MT	OS	Java	V	(Y)
Strabon	1.1	No	3	Adv	N	RDBMS	Y	500MT	OS	Java	N	(Y)
4store	1.1	No	4	No	N	custom	N	15BT	OS	C	V	N
h2rdf+	1.0	No	3	No	N	HBase	N	2.7BT	OS	Java	H&V	N

### C. Datasets of the Smart City RDF Benchmark

The data used for the evaluation are based on the KM4City knowledge base [3]. Some of data are static (or quasi-static) data such as (i) the *road graph* modelling the roads, the public administrations; etc. (ii) the “services” available within the city (e.g., restaurants, hotels, cycle paths, ...) and associated with the road graph and organized in an hierarchy; (iii) the bus stops, bus lines of the local transportation, (iv) the road sensors available on the roads. Moreover, the Km4City model provides a number of hierarchies and structures, and huge data with geolocations in which the inferential aspects of SPARQL queries can be profitably tested. Three different datasets have been adopted for the assessment. They share the same ‘static’ information and only differ for the dynamic part, having one, two or three months of historical dynamic data, respectively. In Table II, the numbers of triples for the different parts of the km4city knowledge base are reported. As you can see, the dynamic parts grow from 22% to 48.5% mostly derived from the AVM (automatic vehicle monitoring, of the ITS) that it is generated out of the data coming for only three bus lines, while the static part is mostly based on structural data like road graph with 34.5M triples, in all the cases.

### D. Real-time data set context description

Since in a real context the dynamic data change regularly (e.g., weather status, AVM, sensors and parking), the behaviour

of the RDF stores should be analysed also under dynamic conditions like queries, while other processes are performing update/upload. Moreover, in order to test a more realistic case the queries retrieving the last value of dynamic data (e.g., sensor last value) could be arranged by using a model including triples stating which is the latest obtained value. In this case, a SPARQL query should be used to remove the association with the latest received value and insert the new triple associated with the new reading of values.

To analyse performance on dynamic update conditions a specific test case has been set up (e.g., traffic, IOT). In order to establish replicable conditions, a tool has been used to regularly generate the status of the 430 sensors using the NTriples format (stored in a specific context) as standard SPARQL 1.1 Graph Store HTTP Protocol. They are produced and singularly loaded into the store, together with their association with the latest value to the corresponding sensor. Each submission stores 19 triples for each sensor and thus 8056 new triples are stored about every 30 seconds. In this case, the 3 months dataset of Table II has been used. Together with the process of upload/update, the server runs at the same time all the queries of the benchmark to assess if updating the triples while querying, either influences or not the query time.



TABLE IV. RDF STORES PERFORMANCE OF DATA LOADING,  
“NA” MEANS THAT THE INFORMATION IS NOT AVAILABLE (IMPOSSIBLE TO MEASURE)

	Triples load time	Stated triples	Stored triples	Size (of which: full text index size, spatial index size)
<i>GraphDB – 1 month</i>	1h 8m	44,274,756	84,425,185	8.5GB (299MB, 66MB)
<i>GraphDB – 2 months</i>	1h 48m	55,617,333	104,041,312	10GB (379MB, 67MB)
<i>GraphDB – 3 months</i>	2h 10m	67,082,202	124,015,329	13GB (459MB, 70MB)
<i>Virtuoso – 1 month</i>	16m	44,274,820	46,259,439	2.2GB (NA, NA)
<i>Virtuoso – 2 months</i>	21m	55,619,789	57,669,629	2.8GB (NA, NA)
<i>Virtuoso – 3 months</i>	31m	67,084,661	69,200,459	3.5GB (NA, NA)
<i>Stardog – 1 month</i>	1h 19m	44,273,368	44,273,368	4.8GB (341MB, 131MB)
<i>Stardog – 2 months</i>	1h 24m	55,615,945	55,615,945	5GB (318MB, 129MB)
<i>Stardog – 3 months</i>	2h 58m	67,080,814	67,080,814	6.2GB (493MB, 138MB)
<i>Oracle – 1 month</i>	6h 18m	44,270,460	78,744,647	25GB (NA, NA)

#### IV. COMPARING RDF STORES WITH SMART CITY RDF ASSESSMENT MODEL

In this section, the RDF stores under assessment are compared according to the feature model which has been identified and discussed in Section III.A. The comparison is carried out with the aim of identifying the stores that are better ranked to be used on smart city applications in terms of provided features.

In Table III, the features supported by the different RDF stores under evaluation are summarized and the values considered as minimum requirements are highlighted. A description of the RDF stores considered in the assessment and reported in Table III is given below.

**Virtuoso 7.2.4** [18], it is mostly known because it is the RDF query engine behind dbpedia.org. It is a SPARQL 1.1 quadruple store developed in C available both via open source and commercial license. The open source version mainly misses the clustering feature. Inference is not materialized at load/indexing time, while query rewrite is performed to support RDFS+ inference. It is backed by the Virtuoso RDBMS and thus SPARQL queries are translated to SQL for that RDBMS. It supports advanced spatial indexing and supports full text search. The community behind virtuoso is headed by OpenLink Software Ltd and it is quite active.

**GraphDB SE 7.0.1** (former OWLIM store) <sup>4</sup> is a commercial solution providing a SPARQL 1.1 endpoint supporting triple/quadruple stores with spatial indexing of geographic coordinates and full text indexing based on Lucene, Apache. It supports inference at load/indexing time with different rule sets (RDFS, OWL2RL, etc.), and such rule sets can be selected by the user. It has been told to support up to 10 billion of triples on a single node. The Enterprise edition allows horizontal scaling with a master node forwarding the insert/update/delete operations to slave nodes. The solution is implemented in Java using OpenRDF Sesame. The project is still active and it is managed by Ontotext.

**Blazegraph** (ex BigData) <sup>5</sup> is an open source project, providing also a commercial license. It supports triple and quadruple stores. With RDFS+ inference (at load time) it is available only on triple stores. It has a full-text indexing support, and there is a basic geospatial indexing, too. It provides both a horizontal and vertical scaling solution, thus allowing an index to be shared on multiple nodes. A single computer can manage up to 50 billion triples. The project is managed by Systap and it is still active.

**CumulusRDF** [19] is an open source project based on OpenRDF Sesame using Apache Cassandra 1.2 as NoSQL storage layer. It does not support inference and can store only triples. Since it is based on Cassandra, it supports vertical scaling for storage of the indexes on the nodes in the cluster, while only one node is used to perform queries.

**Stardog 4.1.1** <sup>6</sup> is a commercial RDF quadruple store developed by Clark&Parsia (developer of the well-known OWL reasoner Pellet). It supports SPARQL 1.1 and OWL2 inference at query time, full-text indexing and search, and spatial indexing and search. It allows horizontal scaling, and it is a quite active project. Stardog may support 10 billion triples store on single node while the community version manages up to 25 million triples.

**Strabon** [20] is an open source SPARQL 1.1 store developed to support both spatial and temporal search [Bereta et al., 2013]. It is based on PostGIS extension of Postgres RDBMS; it does not support inference, nor full-text search. It only provides support for storing triples (the context URI associated with the triple is used for temporal linking). No clustering solution is available.

**4store** <sup>7</sup> is an open source quad RDF store developed in C supporting a clustering solution which stores the quads on different nodes (max 32). It does not support any inference, any full-text search, nor geospatial search. The activity seems to be moved to 5store, which is a corresponding commercial version.

<sup>4</sup> <http://ontotext.com/products/ontotext-graphdb/>

<sup>5</sup> <https://wiki.blazegraph.com>

<sup>6</sup> <http://stardog.com/>

<sup>7</sup> <http://4store.org/>

TABLE V. RDF STORES PERFORMANCE OF NON-SPATIAL QUERIES ( BEST PERFORMANCES IN BOLD)

Query	GraphDB			Virtuoso			StarDog			Oracle	Number of results
	1 month (ms)	2 months (ms)	3 months (ms)	1 month (ms)	2 months (ms)	3 months (ms)	1 month (ms)	2 months (ms)	3 months (ms)	1 month (ms)	
<i>Municipalities-florence</i>	7	10	121	8	15	9	127	173	129	2,391	46
<i>Bus-lines</i>	17	18	91	<b>6</b>	<b>7</b>	<b>6</b>	125	156	141	2,325	85
<i>Bus-stops-of-line</i>	<b>50</b>	<b>26</b>	<b>28</b>	65	68	62	194	211	172	36661	135 (max)
<i>Lines-of-bus-stop</i>	<b>7</b>	<b>12</b>	<b>18</b>	21	23	20	210	235	210	6457	11 (max)
<i>Bus-stop-florence</i>	<b>100</b>	<b>113</b>	<b>126</b>	374	291	281	216	258	201	34071	1108
<i>Bus-stop-forecast</i>	<b>96</b>	<b>413</b>	<b>444</b>	632	2065	2008	2028	3072	5084	259577	15
<i>AVM-distribution</i>	914	1893	2767	<b>26</b>	<b>58</b>	<b>70</b>	1442	2417	3772	26844	89 (max)
<i>Service-florence</i>	7106	7841	10150	<b>2170</b>	<b>2135</b>	<b>2158</b>	3689	3667	3514	>10min	3259
<i>Service-Acc-Clt-Trs-W&amp;F-florence</i>	8,158	8274	8318	<b>2386</b>	<b>2917</b>	<b>2930</b>	4118	4110	6416	>10min	1179
<i>Service-Htl-B&amp;B-florence</i>	3311	3296	4,035	<b>537</b>	<b>845</b>	<b>766</b>	3640	3782	3448	>10min	234
<i>Full-text</i>	314	750	618	<b>64</b>	<b>96</b>	<b>67</b>	166202	214344	215937	136243	1389 (max)
<i>Service-text-florence</i>	286842	295057	284573	<b>1981</b>	<b>3621</b>	<b>5661</b>	165860	202919	209364	126833	51 (max)
<i>Sensor-florence</i>	<b>21</b>	<b>48</b>	<b>46</b>	82	93	84	785	615	483	7349	62
<i>Sensor-status</i>	598	1101	1560	<b>56</b>	<b>146</b>	<b>163</b>	295	384	392	173612	1
<i>Sensor-distribution</i>	939	1867	2665	<b>174</b>	<b>328</b>	<b>341</b>	672	1060	1,346	178272	78 (max)
<i>Parking-status</i>	83	188	309	<b>72</b>	<b>87</b>	<b>100</b>	1,388	1339	1053	40823	1
<i>Parking-distribution</i>	455	1096	1628	<b>61</b>	<b>131</b>	<b>203</b>	223	373	451	30444	83 (max)
<i>Weather-florence</i>	<b>9</b>	<b>19</b>	93	46	60	<b>71</b>	181	182	149	5047	5
<i>Weather-distribution</i>	12	23	19	<b>7</b>	<b>18</b>	<b>11</b>	126	141	128	2342	38 (max)

*h2rdf+* [21] is an open source triple store based on HBase and Hadoop platform. It supports only the SPARQL 1.0 specification, and does not support any inference, any full-text indexing, nor geo-spatial search. Being based on HBase and Hadoop, it provides horizontal and vertical scaling.

*Apache Jena-Fuseki 2.3.1*<sup>8</sup> is an open source SPARQL 1.1 engine integrated within the java based Apache Jena framework. Jena provides the quads RDF storage layer which could be native on file system (TDB), based on a SQL DBMS (SDB) or in memory. Jena provides also the inference support (supporting RDFS, OWL-Lite or using custom rules) but it works only on triple stores and not on quadruples stores, moreover it supports full-text and basic spatial indexing based on Lucene or Solr. No clustering solution has been reported.

**Oracle Database 12c**, the well-known Oracle database solution provides support for RDF graphs, full-text & spatial indexing/search but it does not support the standard SPARQL HTTP query protocol, it can be integrated by using the open source Jena framework with Fuseki or Joseki tools. Moreover

Oracle solution provides inference (RDFS, OWL2RL and custom rules).

As a conclusion of this section, it is self-evident from Table III, that the RDF store solutions supporting all the minimum requirements are Virtuoso 7.2.4 open source and commercial edition, GraphDB Standard Edition 7.2, Stardog 4 and Oracle 12c. Therefore, only these RDF stores have been assessed in term of performance, as reported in Section V.

## V. ASSESSING RDF STORES WITH SMART CITY RDF BENCHMARK

The performance evaluation has been carried out by considering: (i) the loading/indexing time for knowledge base initialization, (ii) the execution time without any update for spatial and non-spatial queries, and (iii) query execution time while the sensors data were regularly updated. The performance has been evaluated using a server Ubuntu 14.04 with 8GB RAM, CPU, Intel Xeon E5-2680@2.8GHz with 20 logical processors, HD at 15.000 RPM. Table IV reports the results for the loading/indexing time concerning the different previously discussed datasets, respectively. It should be remarked that Virtuoso is the fastest, GraphDB and Stardog perform

<sup>8</sup> <https://jena.apache.org>

TABLE VI. RDF STORES PERFORMANCE OF SPATIAL QUERIES (THE BEST PERFORMANCES IN BOLD)

Query	GraphDB			Virtuoso (intersect)			Virtuoso (distance)			Stardog			Number of results
	1 month (ms)	2 months (ms)	3 months(ms)	1 month(ms)	2 months(ms)	3 months(ms)	1 month(ms)	2 months(ms)	3 months(ms)	1 month(ms)	2 months(ms)	3 months(ms)	
<i>Find-address</i>	<b>47</b>	<b>180</b>	<b>218</b>	219	160	143	5762	5965	5776	2495	2848	2367	1
<i>Bus-stop-latlng(100m)</i>	<b>8</b>	<b>8</b>	<b>9</b>	33	63	63	28	31	28	2105	1791	1885	1 (max)
<i>Bus-stop-latlng(500m)</i>	33	52	48	147	182	166	<b>34</b>	<b>41</b>	<b>33</b>	1781	1853	1810	20
<i>Bus-stop-latlng(1km)</i>	82	135	155	76	--	265	<b>42</b>	<b>43</b>	<b>47</b>	2095	2116	2334	93 (max)
<i>Bus-stop-latlng(5km)</i>	463	669	782	--	--	--	<b>201</b>	<b>201</b>	<b>205</b>	2883	3245	2815	1003 (max)
<i>Service-latlng(100m)</i>	691	1111	324	2788	2117	915	<b>582</b>	<b>761</b>	<b>754</b>	3970	4581	5123	41 (max)
<i>Service-latlng(500m)</i>	1131	1256	1212	627	513	549	<b>401</b>	<b>421</b>	<b>391</b>	4110	4303	4467	784 (max)
<i>Service-latlng(2km)</i>	6087	6550	6548	1377	1688	--	<b>1062</b>	<b>1167</b>	<b>1079</b>	5832	6929	6204	3718 (max)
<i>Service-latlng(5km)</i>	11192	13069	12712	3054	--	--	<b>1900</b>	<b>1996</b>	<b>1912</b>	6585	7494	6551	6666 (max)
<i>Service-Acc-Clt-Trs-W&amp;F-latlng(100m)</i>	<b>74</b>	<b>125</b>	<b>87</b>	880	921	773	1260	1209	1073	5978	6076	4986	37 (max)
<i>Service-Acc-Clt-Trs-W&amp;F-latlng(500m)</i>	<b>948</b>	<b>1091</b>	<b>1602</b>	2159	1709	1698	1159	1187	1232	6130	6738	5933	650 (max)
<i>Service-Acc-Clt-Trs-W&amp;F-latlng(2km)</i>	4731	5644	7999	--	--	--	<b>1706</b>	<b>1807</b>	<b>1619</b>	6451	8669	7353	2224 (max)
<i>Service-Acc-Clt-Trs-W&amp;F-latlng(5km)</i>	7983	9610	9974	--	--	--	<b>1785</b>	<b>1938</b>	<b>1813</b>	8024	9669	7646	3102 (max)
<i>Service-Htl-B&amp;B-latlng(100m)</i>	<b>29</b>	<b>38</b>	<b>29</b>	420	466	392	541	563	631	3843	3985	3886	7 (max)
<i>Service-Htl-B&amp;B-latlng(500m)</i>	<b>293</b>	<b>371</b>	<b>393</b>	1119	724	1032	555	666	544	4055	4605	4424	151 (max)
<i>Service-Htl-B&amp;B-latlng(2km)</i>	1390	1729	2020	--	--	--	<b>617</b>	<b>683</b>	<b>681</b>	5664	6573	6278	488
<i>Service-Htl-B&amp;B-latlng(5km)</i>	2421	3144	8281	--	--	--	<b>673</b>	<b>744</b>	<b>682</b>	6516	7009	7053	611(max)
<i>Service-text-latlng(500m)</i>	204936	236937	256328	433	148	324	<b>242</b>	<b>64</b>	<b>73</b>	>7min	>7min	>7min	21 (max)
<i>Sensor-latlng(100m)</i>	<b>10</b>	<b>12</b>	<b>13</b>	--	--	63	20	27	30	1842	1639	1604	0
<i>Sensor-latlng(500m)</i>	41	62	198	118	73	163	<b>18</b>	<b>23</b>	<b>26</b>	1788	2069	1923	4
<i>Sensor-latlng(2km)</i>	229	335	444	--	--	--	<b>22</b>	<b>29</b>	<b>29</b>	2372	2798	2670	29
<i>Sensor-latlng(5km)</i>	514	721	888	--	--	--	<b>23</b>	<b>30</b>	<b>38</b>	2961	2947	2837	56

similarly (about 5 times slower than Virtuoso) and Oracle is the slowest being about twenty three times slower than Virtuoso. Due to the performance of Oracle 12c in loading, the decision was to test only the 1 month data set case. On the other hand, GraphDB and Oracle perform inference at load time while Virtuoso and Stardog at query time, under user request. For this reason, the number of triples indexed by GraphDB is typically 80% bigger than those of Virtuoso. As to Virtuoso, the slight increment of triples stored/indexed with respect to the ones provided to the RDF store (2.1M for the 3 months case) is due the transformation of the geo:lat and geo:long triples in a geo:geometry with POINT() to enable the geo-spatial indexing. While in the same case, as to GraphDB, the increment of about 57M of triples is due to the materialization of triples via inference at the indexing/loading time.

Tables V and VI focus on the results for the query execution time concerning GraphDB, Virtuoso, Stardog and Oracle and related to the different time horizons of one, two and three months, respectively. Table V reports the performances for non-spatial queries and Table VI for spatial queries. The queries have been tested performing a pseudo-random sequence of 1000 queries repeated two times with some pseudo-random arguments in order to reduce the caching effect. The sequence of performed queries has been the same for each test execution, so as to test the same sequence on different systems. The table reports the maximum number of results obtained for each type of query, when the number of results depends on the parameter randomly chosen (e.g., lines of a bus stop) or from the different dataset used (e.g., the AVM, sensor, parking and weather distribution queries). When considering the poor performance by Oracle 12c in loading and also in the query times, it was decided to test only the 1 month



case. Moreover, a bug in the Oracle plugin for Apache Jena did not allow to perform spatial queries via the HTTP protocol and this is the reason why Oracle 12c does not appear in Table VI.

If observing the query results (see Table V), when no spatial and full text search and inference are involved, the performances of Virtuoso and GraphDB are comparable, in some cases GraphDB is even better ranked. When inference is needed (e.g., in the test cases *Service-florence*, *Service-Acc-Clt-Trs-W&F-florence*, *Service-Htl-B&B-florence*), as to Virtuoso the inference had to be enabled on the single constraint involving a general class (e.g., all services in the Accommodation class). While if the inference is enabled, generally on the query, the internal automated query rewrite takes a longer time (may be related to the size of the exploited ontology). For example, for query *Service-Acc-Clt-Trs-W&F-florence* the time grows from an average of 2.9s to an average of 19.6s (on the 3 months dataset). In those cases, the GraphDB results are better ranked. Stardog generally is the slowest on all the queries.

When considering the spatial indexing (see Table VI) in Virtuoso, some mistakes have been detected using the *st\_intersection* function. In some cases, Virtuoso returned an error, in other cases it provided a smaller number of results than the correct number; Virtuoso could provide different results for the same query for the three different datasets, even if they do not differ for the part considered in the query. On the other hand, in Virtuoso, if the *st\_distance* function is used, all the obtained results have been verified to be correct, apart from few cases on the border (due to the numerical computation in measuring distances). The usage of the distance function for Virtuoso is a good solution in most cases, while the query optimizer seems to avoid the exploitation of the spatial index. This fact may be deduced out of a comparison among the results of the formalization of query *Find-address*: in two cases by using: (i) *st\_distance* function it takes about 5.7s, while (ii) with *st\_intersect* function it takes about 0.14s.

TABLE VII. SENSOR DATA UPLOAD PERFORMANCE

	GraphDB	Virtuoso	Stardog
total mean time (ms)	4135.59	1290.05	42498.80
mean upload time (ms)	2105.06	893.52	41526.02
mean update time (ms)	2030.53	396.53	972.78
minimum total time (ms)	1810.00	480.00	6050.00
maximum total time (ms)	37294.00	20678.00	791083.00
std. dev of total time (ms)	2197.81	2082.30	76121.02

TABLE VIII. STORE PERFORMANCE IN PRESENCE AND ABSENCE OF UPDATES DURING BENCHMARK

RDF store	MNQPH		
	no updates	during updates	Loss in performance
GraphDB	2117.00	1799.93	14.97%
Virtuoso	4584.16	4362.21	4.84%
Stardog	1620.24	876.63	45.89%

TABLE IX. PERFORMANCE IN ACCESSING TO THE LAST VALUE OF SENSORS

RDF store	Mean Time to Sensor Status access, Time in ms		
	Case 1 (no update)	Case 2 (no update)	Case 3 (update)
GraphDB	1561	31	334
Virtuoso	163	46	174
Stardog	393	208	554

Another aspect to be considered is the mixing of spatial query with text search query (for example, in query *Service-text-latlng(500m)*). With GraphDB and also with Stardog, we obtained a higher execution time, hitting in some cases the timeout. In this case where spatial and text are combined for Virtuoso, the intersect function returned an error, while the distance function performed very well.

Regarding the analytic queries (for example: *Weather-distribution*, *AVM-distribution*) which count the daily number of records of the weather forecasts, bus, sensor data, parking status for the three datasets, both solutions have provided acceptable execution time (less than 5s). In this case, Virtuoso is better ranked with less than 0.5s of execution time. Moreover, Virtuoso presents a less growing factor with respect to GraphDB.

#### A. Assessing query execution time under update/load

During the test, the time to upload/update new triples for all the sensors and mark them as the ‘latest value’ has been recorded and reported in Table VII. Therefore, the minimum, maximum, average time and the standard deviation of the upload and update time are reported for each RDF store. From the results, it is clear that Virtuoso turned out to be the smartest, since it performed the update of the 430 sensors within 20s, while GraphDB did the same in 37s, and StarDog had an average of 42s and with a maximum time in just one case of 13 minutes to upload new triples for all the 430 sensors. As to Oracle with Apache Jena-Fuseki it was not possible to send the triples for the 430 sensors through Fuseki, since the communication was hanging; while when sending the data for only 10 sensors the average time was about 16s with a maximum of 2.5 min.

In order to evaluate the impact of the update/upload action on query performance, the mean number of query per hour (MNQPH) has been computed for each RDF store in presence or absence of ongoing upload/updates. MNQPH has been computed as the ratio of total time needed to run a large number of queries of the benchmark and the number of queries. In particular, some of the queries such as: “Service-text-latlng(500m)”, “Service-text-florence” and “Full-text” have been removed because they typically generate on GraphDB and StarDog many timeouts, that could create too noise on assessing query performance during update/load activity.

The results are reported in Table VIII, where you can see that the MNQPH is decreased in all the cases, shifting from the value registered with RDF store under no updates up to the value registered during the store updates. The decrease in performance is due to the fact that the query has to wait for the

store unlock. Among the RDF stores considered, Virtuoso presented the lower reduction in performance. Moreover, as stated above, the benchmark occurred in some time outs with to GraphDB and StarDog stores in the absence of updates; typically 46 and 96 times for the whole benchmark. The number of timeouts is more than twice in presence of updates.

Table IX reports the mean time to get access to the latest value of a sensor series (the *Sensor-status* query) in three cases: (1) using the order by clause and without concurrent updates, (2) using the “latest value” triple without concurrent updates and (3) using the latest value triple during concurrent updates. For all the stores we can see that when avoiding the sort and using the “latest value”, the time needed to access is reduced. However, performing a concurrent update increases access time of a significant amount (i.e., more than 10 times for GraphDB, 3.8 times for Virtuoso and 2.7 for StarDog). According to the access mean time values, Virtuoso could perform better than the others in all the cases.

## VI. CONCLUSIONS

The usage of RDF stores to store smart city data is becoming of wide interest for several applications. In this paper we have proposed a *Smart City RDF Assessment Model* for a comparative study about the state of the art on RDF stores according to their main features and in particular on the SPARQL aspects/features. In addition, the *Smart City RDF Benchmark* has been proposed. The benchmark is based on (i) some datasets of triples (that are grounded on Km4City ontological model) accessible from <http://www.disit.org/smartcityrdfbenchmark>, it can be used only for benchmarking purpose; (ii) a set of SPARQL queries declined for different SPARQL constructs. The benchmark has been defined for smart city services to compare results which can be obtained by using different RDF Stores.

The comparison addressed a number of well-known RDF stores such as Virtuoso, GraphDB, StarDog, and Oracle for the performance aspects. As a general consideration about performance, it should be noted that Virtuoso performs better in presence of less selective queries, thus providing a higher number of results. On the contrary, GraphDB performs better when specific results are searched, thus when a smaller number of results are requested.

## ACKNOWLEDGEMENT

The authors would like to say thanks to ONTOTEXT for granting us the access to a trial version of their RDF store. This works has been developed in the framework of Km4City activity for RESOLUTE H2020, and for REPLICATE H2020 (both are European Commission funded International Projects) and for Sii-Mobility Smart City National project (<http://www.sii-mobility.org>).

## REFERENCES

- [1] Mulligan, C.E.A.; Olsson, M., "Architectural implications of smart city business models: an evolutionary perspective," *Com. Magazine, IEEE*, vol.51, no.6, pp.80,85, June 2013
- [2] Welington M. da Silva, Alexandre Alvaro, Gustavo H. R. P. Tomas, Ricardo A. Afonso, Kelvin L. Dias, and Vinicius C. Garcia. 2013. Smart cities software architectures: a survey. In *Proc. of the 28th Annual ACM Symp.on Applied Computing (SAC '13)*. ACM, New York, NY, USA, 1722-1727.
- [3] P. Bellini, M. Benigni, R. Billero, P. Nesi, N. Rauch, "Km4City Ontology Building vs Data Harvesting and Cleaning for Smart-city Services", *International Journal of Visual Language and Computing*, Elsevier, 2014
- [4] Anthopoulos, L.; Fitsilis, P., "Exploring architectural and organizational features in smart cities," in *Advanced Com. Technology (ICACT)*, 16th Int. Conf. on, vol., no., pp.190-195, 16-19 Feb. 2014.
- [5] Zaheer Khan, Ashiq Anjum, and Saad Liaquat Kiani. 2013. Cloud Based Big Data Analytics for Smart Future Cities. *Proc. of the IEEE/ACM 6th Int.Conf. on Utility and Cloud Computing (UCC '13)*. Washington, 381-386.
- [6] Zaheer Khan, Ashiq Anjum, and Saad Liaquat Kiani. 2013. Cloud Based Big Data Analytics for Smart Future Cities. In *Proceedings of the 2013 IEEE/ACM 6th International Conference on Utility and Cloud Computing (UCC '13)*. IEEE Computer Society, Washington, DC, USA, 381-386.
- [7] Y. Guo, Z. Pan, and J. Heflin. "Lubm: A benchmark for owl knowledge base systems". *J. Web Semantics*, 3(2-3):158-182, 2005.
- [8] C. Bizer, A. Schultz. "The Berlin SPARQL Benchmark". *International Journal on Semantic Web & Information Systems*, Vol. 5, Issue 2, Pages 1-24, 2009
- [9] G. Garbis, K. Kyzirakos, M. Koubarakis. "Geographica: A Benchmark for Geospatial RDF Stores". In the 12th Int. Semantic Web Conf. (ISWC). Australia, Oct. 21-25, 2013
- [10] S. Duan, A. Kementsietsidis, K. Srinivas, and O. Udrea. "Apples and oranges: a comparison of RDF benchmarks and real RDF datasets". In *Proc. of the 2011 ACM SIGMOD Int. Conf. on Manag. of data (SIGMOD '11)*. ACM, New York, NY, USA, 145-156, 2011
- [11] Schmidt, Michael, et al. "SP<sup>2</sup>Bench: a SPARQL performance benchmark." *Data Engineering*, 2009. ICDE'09. IEEE 25th International Conference on. IEEE, 2009.
- [12] D. F. Barbieri, D. Braga, S. Ceri, E. Della Valle, and M. Grossniklaus. "C-sparql: Sparql for continuous querying". In *Proc. of WWW*, pages 1061-1062. ACM, 2009.
- [13] J.-P. Calbimonte, O. Corcho, and A. J. G. Gray. "Enabling ontology-based access to streaming data sources" In *ISWC*, pp. 96-111, 2010
- [14] Y. Zhang, M.-D. Pham, O. Corcho and J.-P. Calbimonte. "SRBench: A Streaming RDF/SPARQL Benchmark" In *Proc. of the 11th International Semantic Web Conference ISWC 2012*. Boston, USA, Nov 2012.
- [15] Le-Phuoc, D., Dao-Tran, M., Pham, M. "Linked Stream Data Processing Engines: Facts and Figures" In *International Semantic Web Conference (ISWC 2012)*. Volume 1380, Boston, USA, Springer (2012) 300-312
- [16] Muhammad Intizar Ali, Feng Gao, Alessandra Mileo, "CityBench: A Configurable Benchmark to Evaluate RSP Engines Using Smart City Datasets", 14th Int. Semantic Web Conference, Bethlehem, PA, USA, October 11-15, LNCS, 2015
- [17] P. Bellini, I. Bruno, P. Nesi, N. Rauch, "Graph Databases Methodology and Tool Supporting Index/Store Versioning", publication on *JVLC, Journal of Visual Languages and Computing*, Elsevier, 2015.
- [18] O. Erling and I. Mikhailov. "Virtuoso: RDF Support in a Native RDBMS". In *Semantic Web Information Management*, pages 501-519. Springer, 2009.
- [19] G. Ladwig, A. Harth, "CumulusRDF: Linked data management on nested key-value stores", 7th International Workshop on Scalable Semantic Web Knowledge Base Systems (SSWS 2011), 2011.
- [20] K. Kyzirakos, M. Karpathiotakis and M. Koubarakis. "Strabon: A Semantic Geospatial DBMS". In the 11th International Semantic Web Conference (ISWC 2012), Boston, USA, 11-15 November 2012.
- [21] N. Papailiou, I. Konstantinou, D. Tsoumakos, P. Karras, N. Koziris, "H2RDF+: High-performance distributed joins over large-scale RDF graphs", *Big Data*, 2013 IEEE International Conference on , pp.255,263, 6-9 Oct. 2013