



## A Comparative Analysis of Semantic Web Databases Based on Scalability and Performance

<sup>1</sup>Naveed Ahmed, <sup>2</sup>Ghulam Nabi, <sup>3</sup>Rahmat Ali, <sup>4</sup>Samiullah Bhatti, <sup>5</sup>Dr. Riaz Ahmed Shaik ,  
<sup>6</sup>Mujeeb-Ur-Rehman Jamali, <sup>7</sup>Dr. Samina Rajper, <sup>8</sup>Dr. Shaista Khan

<sup>1</sup>Department of Computer Science, Shah Abdul Latif University Khairpur Mir's, Sindh, Pakistan,  
(ahmednaveed@gmail.com)

<sup>2</sup>Department of Computer Science, Shah Abdul Latif University Khairpur Mir's, Sindh, Pakistan  
(rajperghulamnabi@gmail.com)

<sup>3</sup>Department of Computer Science, Shah Abdul Latif University Khairpur Mir's, Sindh, Pakistan  
(rahmatrajper@yahoo.com)

<sup>4</sup>Department of Computer Science, Shah Abdul Latif University Khairpur Mir's, Sindh, Pakistan  
(sami.khp@gmail.com)

<sup>5</sup>Department of Computer Science, Shah Abdul Latif University Khairpur Mir's, Sindh, Pakistan  
(riaz.shaikh@salu.edu.pk)

<sup>6</sup>Institute of Mathematic & Computer Science, University of Sindh, Jamshoro , Sindh, Pakistan  
mujeebjamali@usindh.edu.pk

<sup>7</sup>Department of Computer Science, Shah Abdul Latif University Khairpur Mir's, Sindh, Pakistan  
(samina.rajper@salu.edu.pk)

<sup>8</sup>Department of Biochemistry ,Shah Abdul Latif University Khairpur Mir's, Sindh, Pakistan  
shaista\_khan787@yahoo.com

### ABSTRACT

Research is rapidly increasing day by day that taken too much efforts in exploring some interesting and some related publications over the internet.as we already know that every data bases have a different architecture that varies the performance in terms of storage architecture and medium. In this research paper we analyzed of two main big data types of Semantic web that is categorized into two types (i) in memory Native (ii) Non-native Non-memory which are disk reside and Non-native is used for services management for instance, SQL, MySQL, and another is Oracle that is just used for storing purpose. Data bases is very important model specially, when any model come into existence. For instance, when we offer for storing purpose of that data then where it should have o store and then definitely it must be access efficiently. The proposed methodology consist test case for data retrieving and query optimization method to analyze performance of databases. When we talk about access data bases from any source then we query them for accessing. LUMB (Lehigh University Benchmark) is being used for testing performance and it cannot be used for storing data. Semantic Web Data (SWD) give a capability in such a way if anybody want to access / encode related data then it can be retrieved efficiently. Our main objective of research we have compared two types of SWD Native store and Non-native store and then we analyzed them.

After analyzing we calculated scalability and performance of each data bases and also set the questionnaire for feedback and how to analysis of Semantic web Databases.

**Key words :** Semantic Web Data Bases (SWD), Ontology, RDF, research study, recommendation, RDF store, Query Optimization

### 1. INTRODUCTION

Semantic Web gives approach tremendously for integrating information from heterogeneous resources and makes computer more intelligent. And it holds massive data in such a manner so that it can efficiently be retrieved. Semantic Web can be divided into three big categories i.e., store data in main memory that is known as Native database i.e., store data permanently on disk in file format that is known as non-native data base. Specially two main challenges of Semantic Web first one, performance means execute various data management tasks in a user interactive time by using less system resources. It is necessary to survey and evaluate of current semantic web data base because developers claim each one of them perform the best [1]. We have focused on two main types (i) Native store and (ii) Non-native store based on their query optimization because every SWD has different query structure. The current management system like SQL, MySQL and Oracle that is being used for storing

purpose. Database is considered as a key tool of any technology whenever we define about two models that is related to the object or XML data model is offered storage, query and management solution that is directly being implemented in academic and as well as industrial DBMS (Database Management System) this is much needed to manage and to store them in such a manner and that is necessary to keep marketplace. And such kind of technology answered to essential and opened determined result as significance a new type of data is said to be Semantic Databases (SDB) [2]. Georgia Troullinou *et al.* A best valid RDF document / graph defines excellent concept of schema adopting to the corresponding instance [3]. We have proposed some key SWD method and their result. And how databases have been used are described in detail.

Data bases is needed to implement efficiently and it is important for SWD to give some best method for accessing RDF (Resource Description Form) graph. There are many open source are available for checking the performance of SWD that are being used for

1. checking the evaluation performance because these tools are necessary for calculating performance of updated SWD against existence web and it gives facility to choose from library application. We will further discuss about SWD regarding results in detail.
2. To measure scalability and performance we collected 10 databases developer and 25 databases administrator for giving feedback so that we can compare each databases otherwise it was not easy for calculating performance and scalability. Administrator faces many hindrance because administrator have to manage all these important operation like mange databases, backup archive.
3. As we have discussed about Semantic Web and its two main types. By dividing these databases means Native store in which two main other databases come into this kind of SWD first, Virtuoso and AllegroGraph. In Virtuoso SWD we have chosen four actions for checking their performance (open instance, create table, impor.mdfdatabase, create connection, drop table) after checking performance some action was working tremendous but some did not perform well. Some other actions were used in AllegroGraph web database that is Native store in which we has given some parameters like
4. import.mdfdatabase and create connection and some other but above two mentioned parameters that was executed by high optimized just because of RDF and XML. AllegroGraph is a vital for supporting XML schema type. Virtuoso semantic web does not have such capability. AllegroGraph has been designed such a type in which which query speed is much faster than Virtuoso. AllegroGraph is considered as good tool in performance point of view. And that is used as Benchmark RDF file and it has a capability of retrieving approximately 40,000 triples in just a few

seconds. Now we highlight on Non-native SW means SQL web database it demonstrates the capability of performance of SQL database (DDL Data Definition Language). 11 actions have been tested in Non-native SW (*open instance, create table, import.mdfdatabase, create, connection, drop table, deploy database, create backup, create mirror, role distribution, create trigger, passing flag*). Other side AllegoGraph it allows superb for storing triples in terms of data size.

## 2 RELATED WORK

This section describes different tools that is being used to calculate performance of SWD so for LUMB (Lehigh University benchmark) is also a good tool for testing OWL. Effective tool is very less in evaluation point of view analysis. And test case is always presented for managing data techniques and SPARQL (Protocol and RDF Query language) [2]. Test case was developed for knowing performance and scalability of SWD with respect of CRUD operation (create, read, update, delete). Most of two methods SW data loading bulk means loading for incrementing loading and to display store performance [1]. Xianwei shen, *et al.* [4]. Jena frame work plays a vital role in semantic web that always maintains RDF in memory in form of hash map based structure. The performance study that is always usually less of comparison over various RDF (resource Description Framework) data stores with specific storage and query methods. Semantic web performance is considers as vital drawback. Hash map always consist of subject, properties and object for storing three times in GraphTriplesStore. Semantic web works tremendous on formal ontologies because of their structure model is basic data structure for better understanding and transporting specially for machine understanding. Main vision of semantic web is to propose such an environment for data on the web can be semantically interpreted and processed by machine so that it will be also be helpful for human understanding [5, 15]. Test queries have been done by using LUMB tool. Input size that is not considered as class instances. By using LUMB used Metrics set for performance metrics including load time, repository size, query response etc. that is totally benchmark data base metrics. Query Response Time was considered that was used for the process used in benchmark in which query was executed and also calculated the average time of it. In query and soundness was examined it is so necessary with respect to query because the required answer that are always entitled by the knowledge base system (KBS). It was calculated the degree of each and every query on the basis of percentage of required answer which are taken by the system [6].

## 3 PROPOSED EVALUATION METHODOLOGY

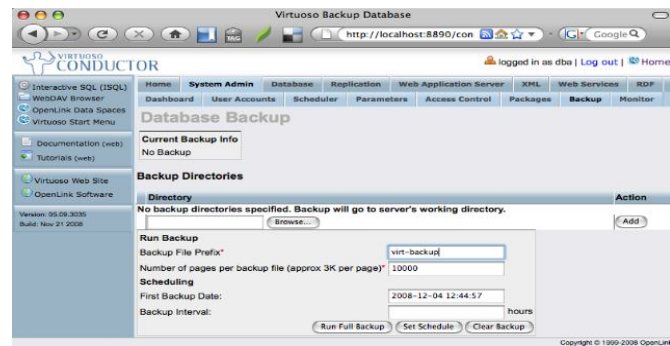
In this section some highlight on evaluation methodology in which we have described some key factors about performance and scalability, detail of all action / queries.

### 3.1 Test Case for Virtuoso web database (Native Store)

Virtuoso web database that is considered as (Native store) in which we have given 5 actions from which only two actions did not perform well as shown in Table.1 as we earlier said that data on the web which is consist of millions of triples (Subject, Predicate, Object). Import.mdfdatabase’s status is zero reason behind that is because it is not easy for few databases server resulting server does not work better. Orri *et al*, SPARQ and RDF specially work on by mapping on demand [7]. It is very difficult for giving table and column cardinalities for RDF triples OR you can say quad table. SQL compiler hosting the as process same that is known as important part of an index are unchangeable at compile time. SPARQL to be prefixed with SPARQL keyword for differentiating it from SQL that is ending point for HTTP is equally present [8].

**Table. 1** Test Case for Virtuoso web database (Native Store)

Case	Actions	Expected Result	Received Result	Status
1	Open instance	Open instance	Open instance	Successful
2	Create table	Table created	Table created	Successful
3	Import.mdfdatabase	Import.mdfdatabase	Import.mdfdatabase	Successful
4	Create connection	Create connection	Create connection	Successful
5	Drop table	Drop table	Drop table	successful
6	Deploy Database	Deploy Database	Deploy Database	Un-successful
7	Creating Backup	Creating Backup	Creating Backup	Successful
8	Create Mirror	Create Mirror	Create Mirror	Un-successful
			Calculated Average	75.0%



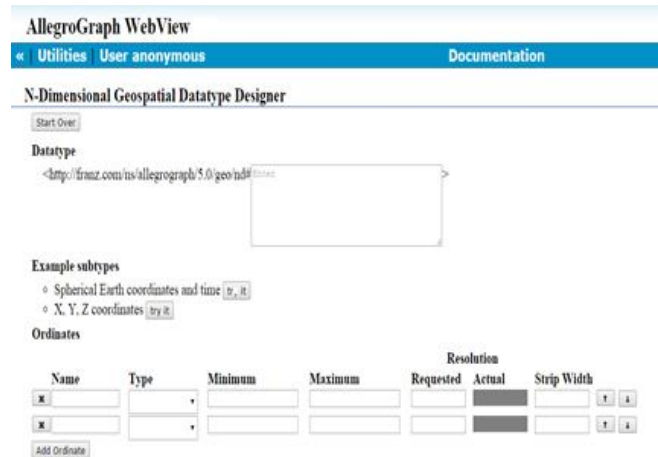
**Figure 1** below is the interface of Virtuoso web

### 3.2 Test case for AllegroGraph web data base (Native store)

Unstructured of system is provided by AllegroGraph that plays important role for server and its triples stores. It is very useful for data base administrator for security purpose and performance. In virtuoso SWD import.mdfdatabase did not perform well but when same query was given to AllegroGraph Web data it gave the result successful and accuracy was 100% correct. As we defined earlier that Virtuoso SWD did not perform well just because of database server does not work smoothly. AllegroGraph uses some extra features that loads RDF data into repository and configure triples for that repository. You can see in Table 2.

**Table 2** (Test case for AllegroGraph web data base (Native store))

Case	Actions	Expected Result	Received Result	Status
1	Open instance	Open instance	100%	Successful
2	Create table	Table created	100%	Successful
3	Import.mdfdatabase	Import.mdfdatabase	100%	Successful
4	Create connection	Create connection	100%	Successful
5	Drop table	Drop table	100%	successful
6	Deploy Database	Deploy Database	0%	Un-successful
7	Creating Backup	Creating Backup	100%	Successful
8	Create Mirror	Create Mirror	0%	Un-successful
			Calculated Average	75.0%



**Figure 2** Interface of AllegroGraph SW view

AllegroGraph that retrieves approximately 40,000 triples in just few seconds for disc no need of serialization and deserialization.

### 3.3 SQL database (non-native store)

SQL web database is estimated, accessed huge number of information and very large number of websites are assisted by SQL database. This semantic web SQL perform tremendously as compared to both previous SWD Virtuoso and AllegroGraph

SW. in SQL just on action did not perform successfully because SQL is not open source and it works on independent. As you can see in Table 3.

**Table 3** SQL database (non-native store)

Case	Actions	Expected Result	Received Result	Status
1	Open instance	Open instance	100%	Successful
2	Create table	Table created	100%	Successful
3	Import.mdf database	Import.mdf database	100%	Successful
4	Create connection	Create connection	100%	Successful
5	Drop table	Drop table	100%	successful
6	Deploy Database	Deploy Database	100%	Successful
7	Creating Backup	Creating Backup	100%	Successful
8	Create Mirror	Create Mirror	100%	Successful
9	Role Distribution	Role Distribution	100%	Successful
10	Create Trigger	Create Trigger	100%	Successful
11	Passing Flag	Passing	0%	Un-successful
			Calculated Average	91.0%

**3.4 Oracle web database (non-native)**

This semantic web is considered as best SWD. Oracle semantic web is an open source and independent database modeling developing. Ontology assisted query and function expand ability query on semantic web. It is a bulk that is awesome concept. All 12 actions executed in this oracle semantic web. It is one of best SW database as than all mentioned above SW database. Oracle is most elastic and rate action process to manage information and application. As shown in Table 4.

**Table 4** (Oracle web database (non-native))

Case	Actions	Expected Result	Received Result	Status
1	Open instance	Open instance	100%	Successful
2	Create table	Table created	100%	Successful
3	Import.mdf database	Import.mdf database	100%	Successful
4	Create connection	Create connection	100%	Successful
5	Drop table	Drop table	100%	successful
6	Deploy Database	Deploy Database	100%	Successful
7	Creating Backup	Creating Backup	100%	Successful
8	Create Mirror	Create Mirror	100%	Successful
9	Role Distribution	Role Distribution	100%	Successful
10	Create Trigger	Create Trigger	100%	Successful
11	Passing Flag	Passing flag	100%	Successful
12	Create form	Create form	100%	Successful
			Calculated Average	100.0%

**3.5 MySQL web database (non-native store)**

MySQL has not proved and acceptable for purpose of triples store. MySQL does not support import.mdf database action resulting MySQL is not better than SQL database because MySQL sever is hundreds of mission critical and heavy load production and deployment software. You can see the result of MySQL in Table 5.

**Table 5** MySQL web database non-native store

Case	Actions	Expected Result	Received Result	Status
1	Open instance	Open instance	100%	Successful
2	Create table	Table created	100%	Successful
3	Import.mdf database	Import.mdf database	0%	Un-Successful
4	Create connection	Create connection	100%	Successful
5	Drop table	Drop table	100%	successful
6	Deploy Database	Deploy Database	100%	Successful
7	Creating Backup	Creating Backup	100%	Successful
8	Create Mirror	Create Mirror	0%	Un-Successful
9	Role Distribution	Role Distribution	100%	Successful
10	Create Trigger	Create Trigger	100%	Successful
11	Passing Flag	Passing	100%	Successful
			Calculated Average	81.0%

Following table 6 highlight on all comparative results of SWD which is calculated separately after receiving result we have mentioned separately each semantic web data base. For a web scale infrastructure system should be methods from entire the data base spectrum [14].

**Table 6** (comparative results of all semantic web data bases)

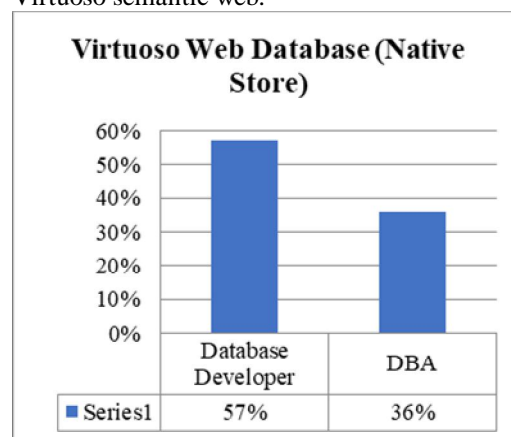
S. no	Semantic Web Databases Name	Calculated AVG:
1	Virtuoso Semantic web Database (Native store)	60.0%
2	AllegroGraph Web Database (Native store)	75.0%
3	SQL database (Non-native store)	91.0%
4	MySQL database (Non-native store)	81.0%
5	Oracle database (Non-native)	100.0%

### 3 RESULTS AND DISCUSSIONS

This section consists of all semantic data bases that was produced results by us in terms of analysis and also for evaluation that we have gain throughout our research.

#### 3.1 Evaluation results of Virtuoso web database for performance

Generating the results of Virtuoso SWD that is kind of native store. In this semantic web data developer gave positive result because it has better features of storing and retrieving functions. And their performance is also awesome in terms of retrieving whereas data base administrator is behind data developer because that is not suitable for database administrator. Resulting 25% more better result than database administrator. Alexander Potocki *et al.* native store has awesome ability of storing RDF progress and this is closing performance ability for RDMS [9]. Andre B.bondi *et al.* when we discussed about scalability that is very important factor and it is crucial for it especially for long time success and bad system performance [10]. Following figure 3 shows the evaluation result for Virtuoso semantic web.



**Figure 3** (evaluation result of Virtuoso SW)

#### 3.6 Evaluation results of AllegroGraph web database for performance

We got the result for AllegroGraph higher for data base developer and not better for results seen for data base administrator because main object of data base developer is to design, test and implement existing data base. While data base administrator just uses the specialized software for storing and organizing data. AllegroGraph also works on RDF store and better performance and it always suitable for GUI (graphical user interface) better for exploring, querying and maintain AllegroGraph triple store. AnHai Doan *et al.* Now a days data bases are being used widely and it is very important to interpret between multiple databases [11]. AllegroGraph data base that is kind of native store we received result 65% for data base developer and 41% for data base administrator. We can say that

AllegroGraph is reliable for data base developer whereas for DBA it is not better for performance point of view. Following figure 4 shows the evaluation result for AllegroGraph semantic web.

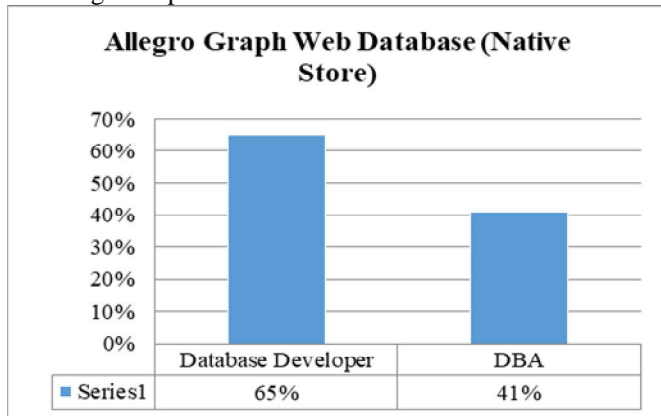


Figure 4 (evaluation result for AllegroGraph SW)

### 3.7 Evaluation result of SQL web data base for performance

SQL web data base has a key features which comes into category of non-native store as native store does not care for it. And it is different SW data base because it has lot of tables and also better for store process and methods and gives power for trigger. Like parsing parameter that gives function trigger.in SQL DBA is ahead resulting 81% while data base developer 77%. Figure 5 shows the evaluation result for SQL SW.

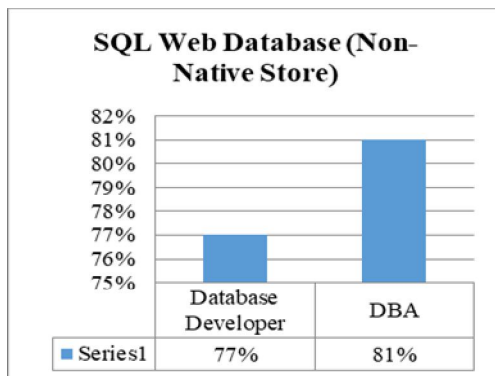


Figure 5 (evaluation result for SQL SW)

### 3.8 Evaluation result of MySQL web database for performance

Performance of MySQL is higher than DBA, MySQL which is type of non-native semantic web store. We received the results for that superb 91% for data base developer whereas 70% for DBA. Some since some decade MySQL is being used widely in terms of user. We know that it always works on client-side server and it is much helpful for data warehousing, e-commerce and some logging application. You can see the figure 6.

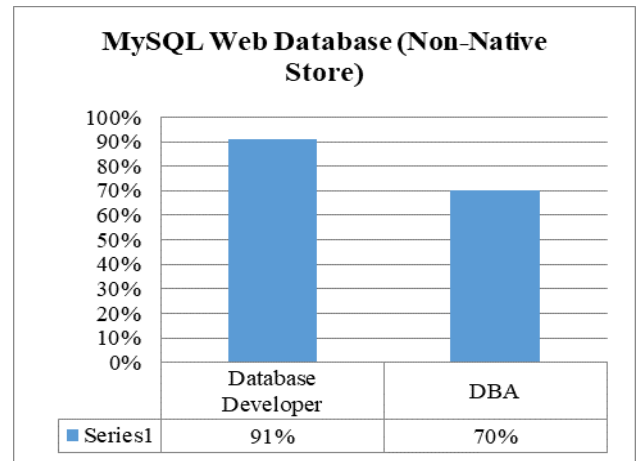


Figure 6 (evaluation result for MySQL SW)

### 3.9 Evaluation result of oracle we data base for performance

Oracle is tremendous choice for any organization it is considered as best language. It totally open-source and much used by DBA. It works on RDF, RDFS and ontologies by jena, sesame, etc. Oracle gives the security like virtual private data base and for semantic indexing and for document which is store in oracle data base. It has superb capability of storing trillions of triples that is much need of today’s semantic web data base. You can see the result in figure 7

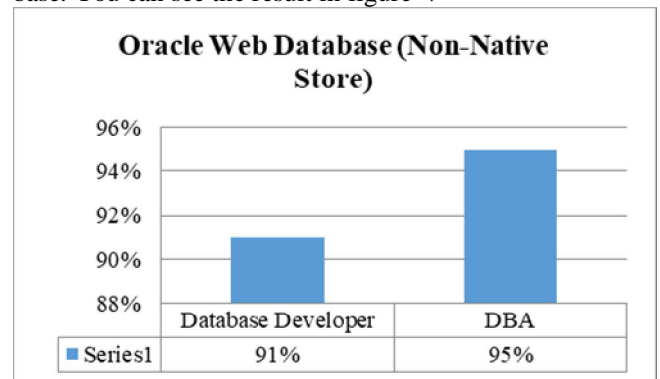


Figure 7 (evaluation result for Oracle SW)

Table 7 shows the Evaluation Results for Database Developer and Database Administrator (DBA) in terms of Scalability and performance

Database Name	Database Developer	DBA
Virtuoso web database (native store)	57%	36%
AllegroGraph web database (native store)	65%	41%
SQL web database (non-native store)	77%	88%
MySQL web database (non-native)	91%	70%
Oracle web database (non-native)	91%	95%

#### 4 CONCLUSION AND FUTURE WORK

It is very important to develop a data campaign for fair and realistic performance. As we know that new important research that will be helpful for semantic web community. And a lot work has been done so far and still work to query performance and optimization and large numbers of publications have been published so far. And another technique is to mining data and for extracting information for knowledge effective point of view. Cliff Joslyn *et al.* Now a days triples are semantic web increasing rapidly to store SPARQL over SQL architecture data base and it is helpful to understand their semantic structure [12]. Deepika Chaudhary *et al.* the semantic web is robust and sensible approach over multitude of data. The semantic web gives a platform to design web pages that can be understood easily by computer [13]. The MCQS questionnaire was designed and distributes among the students and faculty members of Computer Science and also other platform like social network and google.docs. The accuracy of native store databases was recorded 61% and 38.5% was seen result for DBA. Here 86.33% for data base developers and 88.5% for DBA. The evaluator which is working on professional organizations including Kalpoint Veripark, QuadQ were fully satisfied from oracle data base is one of the best data base that is supported platform independent. The scalability and performance was measured via the test case and received results in some actions were performed brilliantly and some did not response better in terms of performance. We had selected five databases from which two databases were selected from native store and three databases from non-native store. Finally we had got tremendous result about oracle data base in which we had seen tremendous flexibility in terms of query and storing. Oracle is one of the best database. This research is highly beneficial for fYP students of BS (CS) or (IT) while choosing their data bases for their projects.

#### REFERENCES

- [1] Butt, Anila Sahar, and Sharifullah Khan. "Scalability and performance evaluation of semantic web databases." *Arabian Journal for Science and Engineering* 39, no. 3 (2014): 1805-1823.
- [2] Guo, Yuanbo, Zhengxiang Pan, and Jeff Heflin. "LUBM: A benchmark for OWL knowledge base systems." *Journal of Web Semantics* 3, no. 2-3 (2005): 158-182.
- [3] Troullinou, Georgia, Haridimos Kondylakis, Evangelia Daskalaki, and Dimitris Plexousakis. "RDF digest: Efficient summarization of RDF/S KBs." In *European Semantic Web Conference*, pp. 119-134. Springer, Cham, 2015.
- [4] Shen, Xianwei, and Vincent Huang. "A framework for performance study of semantic databases." In *IWEST@ ISWC*. 2010.
- [5] Giri, Kaushal. "Role of ontology in semantic web." *DESIDOC Journal of Library & Information Technology* 31, no. 2 (2011).
- [6] Chakkarwar, V. A., and Amruta A. Joshi. "Semantic web mining using RDF data." *International Journal of Computer Applications* 975 (2016): 8887.
- [7] Erling, Orri, and Ivan Mikhailov. "RDF Support in the Virtuoso DBMS." In *Networked Knowledge-Networked Media*, pp. 7-24. Springer, Berlin, Heidelberg, 2009.
- [8] Erling, Orri, and Ivan Mikhailov. "RDF Support in the Virtuoso DBMS." In *Networked Knowledge-Networked Media*, pp. 7-24. Springer, Berlin, Heidelberg, 2009.
- [9] Potocki, Alexander, Anton Polukhin, Grigory Drobyazko, Daniel Hladky, Victor Klintsov, and Jörg Unbehauen. "OntoQuad: native high-speed RDF DBMS for semantic web." In *International Conference on Knowledge Engineering and the Semantic Web*, pp. 117-131. Springer, Berlin, Heidelberg, 2013.
- [10] Bondi, André B. "Characteristics of scalability and their impact on performance." In *Proceedings of the 2nd international workshop on Software and performance*, pp. 195-203. 2000.
- [11] Doan, AnHai, and Alon Y. Halevy. "Semantic integration research in the database community: A brief survey." *AI magazine* 26, no. 1 (2005): 83-83.
- [12] Joslyn, Cliff, Bob Adolf, Sinan al-Saffar, John Feo, Eric Goodman, David Haglin, Greg Mackey, and David Mizell. "High performance descriptive semantic analysis of semantic graph databases." In *1st Workshop on High-Performance Computing for the Semantic Web at ESWC*. 2011.
- [13] Deepika Chaudhary, Jaiteg Singh, D. P. Kothari. "Performance Evaluation of the Semantic Web Reasoners". *International Journal of Recent Technology and Engineering (IJRTE)* ISSN: 2277-3878, Volume-8 Issue-4, November 2019.
- [14] Erling, Orri, and Ivan Mikhailov. "Towards web scale RDF." *Proc. SSWS* (2008)
- [15] Mujeeb-ur-Rehman Jamali, Abdul Ghafoor Memon, Nadeem A. Kanasro, Mujeeb-u-Rehman Maree, (2021). "Data integrity issues and challenges in next generation non-relational document-oriented database outsourced in public cloud", *International Journal of Emerging Trends in Engineering Research*, Volume 9. No. 4, April 2021, pp-416-420. ISSN 2347-3983. <https://doi.org/10.30534/ijeter/2021/13942021>.