International Journal of Engineering & Technology, 7 (2.33) (2018) 287-289

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

A knowledge-based stream processing using big data analytics

D. Tamilselvi [1] *, Dr. A. Akila [2]

[1] M. phil Scholar, Department of School of computing Sciences ,VISTAS , Chennai, India
[2] Assistant Professor, Department of School of computing Sciences, VISTAS, Chennai, India
**Corresponding author E-mail: tamil3101@gmail.com*

Abstract

Big Data available in almost all departments of every organization spread throughout the globe in a huge volume and category. The issues such as data heterogeneity and advanced processing capabilities are provided solution with the proposed system. Data heterogeneity is tack-led by using the automatic schema mapping in the proposed work which is knowledge based solution. Inventive processing is achieved using ontology extraction and semantic inference in the proposed work. The solution is evaluated in terms of its performance and effective-ness with the publish/subscribe paradigm. The state of art analysis of huge volume of variety of data and sensory information is the real complexity. The Advanced Message Queuing Protocol is used in the proposed work for the state of art substance explanation of flooding IoT data to have dynamic mingling. The proposed work gives way to produce huge amount of data that can affect the working of the smart city systems that uses IoT data. To point out the reliability and summarization of the data, information model is used in the proposed work. The data size and average exchanged message time are the measures used to examine the working of the framework. A detailed assessment of the various sensors is carried out to inspect the storage data volume and computational cost for the substance explanation of the framework.

Keywords: *Subscribe Services; Interoperability Schema Matching Semantic; Big Data Analytics Ontologies*

1. Introduction

The connection between the different items or stuffs is described by IoT that allows verification of the actual globe thing and providing the virtual versions of the internet. The availability of these linked substances and sensors provides path to know the status of the current real world and make use of this information to raise the competence of the city services and infrastructure. In smart cities, knowledge management is a main point and the solution available till now provides solution for different circumstances and applications are in remoteness. The information about traffic available in Wikipedia and data that foretell about weather does not coordinate with each other even though they work with same kind of data which leads to a situation where the information available in one resource is not use information available in another resource. The effect of this problem could be apparently proved once the exchange of results between IoT researchers is taken in to consideration. Combining the outcome with other data sources will permit them to be useful to others.

Data Integration: With the available cities are alive and enhancing eco- systems, it is necessary to have a nonstop connection and realize the dynamic knowledge between city stakeholders and citizens to make use of information it gets expired. The current stream justification is yet another serious problem that should be tackled with in smart city framework. To provide solution to the problems within the area of smart cities, the proposed work describes a solution known as City Pulse for present IoT stream explanation that uses a knowledge-based method to depict data sets using Mashup platform. The framework also gives an information model to give a representation for summarization and trustworthiness of IoT stream data.

Clustering analysis is an emerging research area with lot of application areas. For the past few years, many clustering algorithms are available used in variety of application areas such as image processing, manipulation product, mobile product, medicine and economics. Not being identical is the major trouble with data clustering algorithms. So we will go to JC8 algorithm.

Clustering Algorithm Process: Algorithms are designed may provide better solution with one category of data set, but may be unsuccessful. It may give bad result with data set of other types. Making a homogenize algorithm suitable for all situations may take many trials, but the major requirement has not been met. The data set is segregated into homogeneous sets based on the characteristics to assemble similar objects in same group and dissimilar objects are in distinct groups.

J48 Algorithm Process: The solution provided by in proposed work showcases the efficiency and accuracy of J48. Data mining could be implied in many category of information repository. To get data from product and implement to SQL server in back end connections.

Methodology: This research proposal is at the crossing of middleware for large-scale platforms working on large data volume, languages, big data, social networking. The distributed product year, number of items, cost details and publish allowed her to gain experience in situational-awareness through complex event processing and supporting publish/subscribe platforms for web-semantic described events.

Experimental Setup: These processes permits one to describe a program as taking ultimately, after a compilation process, the form of a DAG (directed acyclic graph) but according to our under-

Copyright © 2018 D. Tamilselvi, Dr. A. Akila. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

standing, none permits to become accustomed with the program at runtime with respect to its functional/business nature. As an outcome, developers concentrate on conveying data processing logic, cost, product items and product count but not coordinating runtime variation. Few data workflow engines initiates to be viewable in the group of large-scale stream processing but without yet an required behavioral/functional accustomed capability, as the so far only attention was to permit usability to experience necessities for non-functional modification measured by scalability, fault-tolerance, performance needs through node replication or elasticity.

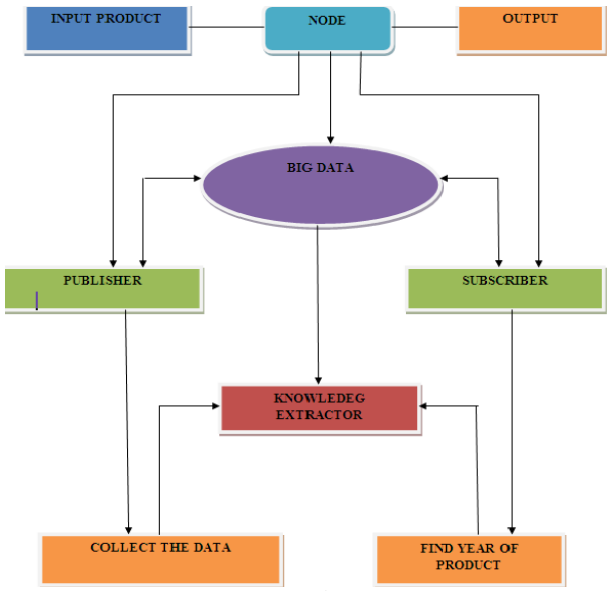


Fig. 1: The Workflow Diagram.

2. Literature review

2.1. Semantic coordination

A New Approach and an Application
The work described in this paper is only the starting phase of a very aspiring scientific challenge, such as examining the least general ground required to make the communication between independent entities that cannot view into each other's head, and thus can attain some amount of semantic synchronization only through other means, like exchanging examples, focusing to things, recalling past interactions, making it common from past communications, and so on. To this end, a lot of work residue to be done. We show how to implement the model and the algorithm to a motivating family of semantic models, namely hierarchical classifications, and give the outcome of initial tests on two categories of hierarchical classifications, Web directories and catalogs. At last, we give the considerable development on earlier approaches.
Fusion of Geospatial Perceptive, Temporal Logic, & Societal Network Exploration in an RDF Database: Many people were engaged to analyze the result of the proposed framework. The approach describes the capabilities that can be implemented in one query language, in this case Prolog. And we assume that in the recent future, these facilities will exist. A social network database along with its events in time is of larger volume when compared with the email and calendar database of an large organization. An event can have many additional attributes and annotations. On top of this event database we implemented libraries for RDFS++ logic reasoning, for geospatial and temporal capabilities, and an extensive social network analysis package.

2.2. EP-sparql

The Event Processing and Stream Reasoning technologies are allowed to handle knowledge management in terms of ontology mechanism which is represent slower knowledge development and time invariant. The main aim of this proposed work is to provide a basic structure of Event Processing and Stream Reasoning exceeding modern technology for in both Event Processing and Semantic Web. The semantic web is to handle knowledge management and Stream reasoning, but cannot be able to deal with changing the data which is provided by event streams. The main concern of Event Processing is timely detection of compound events within the data streams of simple events. The open-source technique provides such mechanism to implement and present a combination of results to shown the effective and convenience of proposed work.
The main aim of this framework is based on publisher and subscriber services directly interconnect with the big data analytics and stream processing. It is likely to come across some difficulty because of the various types of schema recognized by the heritage structures to be overstated. Related to these issues, the successful notification of events followed by the subscribers used different kind of schemas rather than publishers. It is also experimentally confirmed that the solutions are capable to publisher/subscriber service based on data schema, which latency degradation is affected by some flexibility. This framework is to apply for data heterogeneity relate with face recognition based on the automatic schema mapping technique and also ontology extraction and semantic inference support to advanced processing.

3. Data collection

The enterprise product data is collected for many products. The application of enterprises data mining includes the availability of objects which are connected from the nodes and wireless sensor nodes to monitor the current status of the real world in real time. The process information is to improve the efficiency of operational performance of city services and the infrastructure of the environment.

3.1. Data set working process

The connectivity based clustering is a entire product of methods that vary by the method sales are manipulated. Separately from the classic choice of distance product, the user also needs to determine on the linkage criterion to use. Popular choices are identified as multiple products to promise and publishing date will be stated in the dataset collection. Gathering the project items and number of products is to combine with all product items.

Table.1: Data Set

A	B	C	D	E	F	G	H	I
Name	Gender	DOB	Class	Subjects	Health related problems	Father's occupation	Affected by any financial problem	Income
K.Chandersekn	female	10.03.1997	12	chemistry	yes	self Employee	Family level	25,000
R.Dhana	male	07.05.1996	12	chemistry	no	self Employee	no	10,000
S.JayaBharathi	female	18.05.1996	12	chemistry	no	self Employee	no	50,000
J.Kannan	male	20.06.1998	12	chemistry	no	self Employee	no	15,000
S.Nalina Santha Kumari	female	04.11.1992	12	chemistry	no	self Employee	no	25,000
S.Nandhini	female	01.05.1996	12	Maths	yes	self Employee	no	25,000
G.Rajesh	male	08.06.1997	12	Maths	no	self Employee	no	50,000
M.Ramya	female	08.06.1997	12	Maths	no	self Employee	Edictional level	50,000
R.Santhya	female	12.05.1997	12	Maths	no	self Employee	no	10,000
R.Sarath	male	11.06.1991	12	chemistry	no	self Employee	no	50,000
E.Tamilselvi	female	10.06.1996	12	chemistry	yes	self Employee	no	15,000
V.Vennila	female	07.05.1995	12	chemistry	no	self Employee	no	25,000
S.Vinoth	male	20.05.1997	12	chemistry	yes	self Employee	no	25,000
A.Vimal	male	24.09.1996	12	chemistry	no	self Employee	no	50,000
S.Anitha	female	12.07.1998	12	chemistry	no	Employee	no	50,000
R.Bharathi	female	04.03.1997	12	chemistry	no	Employee	no	50,000
P.Kalpana	female	24.09.1997	12	chemistry	no	self Employee	no	50,000
M.Meena	female	12.06.1998	12	chemistry	yes	Unemployee	Family level	10,000
S.Ram	male	11.05.1998	12	chemistry	no	self Employee	no	50,000
T.kavin	male	05.06.1998	12	chemistry	yes	self Employee	no	15,000
S.Babu	male	30.06.1997	12	Tamil	yes	self Employee	no	25,000
S.Iluvarasi	female	24.07.1997	12	Tamil	yes	self Employee	no	25,000
P.Keerthana	female	12.04.1997	12	Tamil	yes	self Employee	no	50,000

The information's of all product items and query eloquence i.e, actions composing the data streams to be collect and consider as a attributes and values of map, and the queries on event streams are organized to extract the values in the events. The data dispensation required by multipart Event dispensation is characteristically recognized values based on the computing rules shown by some attributes of the received events. Generally, tree-based formalism with two different set of nodes are classified, one for entity representation and another for their dataset attributes.

3.2. SQL server data

Table.2: SQL Server Data

G015\SQLEXPRE...- dbo.data_set | Summary

Column Name	Data Type	Allow Nulls
productname	nvarchar(50)	☐
productno	nvarchar(50)	☐
cost	nvarchar(50)	☐
noofitem	nvarchar(50)	☐
subscribe	nvarchar(50)	☐
publish	nvarchar(50)	☐
reviews	nvarchar(50)	☐

3.3. Related work

Data workflow engines starts to execute in the community of huge stream dealing out but without yet an sufficient behavioural/functional alteration capability, as the so far only concentration was to permit alteration of information to look necessities for non-functional modifications done by product item, cost, number of product and year performance requires through node replication or elasticity. The literature survey carried out so far only offers a solution for functional adaptively in stream processing languages of big data platforms, even if some fore works on stream platforms or active databases are a good starting point.

4. Conclusion

The advanced message queuing protocol is used in the proposed stream annotation framework for current SQL stream to help to provide huge volumes of data. This framework is experimented with various features of data stream, raw data and aggregated data, hence easily identify the location to improve the working model of our annotated data. The evaluation metrics are size of the data and average time of message exchange exhausted over the communication which helps to manipulate the proposed work. It was established in all category of performance framework which is improved as 99.4% and 96.2% with annotated summarized data stream with respected to the size of the data and average exchange time. The future work will be implement a huge set of data stream and make use of a network with real-time road traffic to evaluate working of the proposed work with huge number of clients in detail. In future, this work is needed for data stream analysis techniques which are used by the research scholar, when compared to other existing research tools like generalization and harmonization model. This proposed work is wide range of vocabulary to wrap representation of data analytics features formed by modern stream analysis techniques.

Acknowledgement

I would like thank Vels Institute of Science, Technology & Advanced Studies, Pallavaram for their constant support to carry out my research.

References

[1] K. Teymourian, O. Streibel, A. Paschke, R. Alnemr, C. Meinel, Towards semantic event-driven systems, in: Proceedings of the 3rd International Conference on New Technologies, Mobility and Security (NTMS), 2009, pp. 1–6.

[2] F. Heintz, J. Kvarnström, P. Doherty, and Stream reasoning in DyKnow: a knowledge processing middleware system, in: Proceedings of the Stream Reasoning Workshop. In series: CEUR Workshop Proceedings.

[3] K. Teymourian, O. Streibel, A. Paschke, R. Alnemr, C. Meinel, Towards semantic event-driven systems, in: Proceedings of the 3rd International Conference on New Technologies, Mobility and Security (NTMS), 2009, pp. 1–6.

[4] D. Brickley, R.V. Guha, RDF Vocabulary Description Language 1.0: RDF Schema. <http://www.w3.org/TR/rdf-schema/> (accessed July 2013).

[5] E. Prud'hommeaux, A. Seaborne, SPARQL Query Language for RDF. http:// www.w3.org/TR/rdf-sparql-query/ (accessed July 2013).

[6] D. Crockford, the Application/json Media Type for JavaScript Object Notation (JSON), RFC4627. <http://www.json.org/index.html> (accessed September 2013).

[7] O. Ben-Kiki, C. Evans, YAML Ain't Markup Language. <http://www.yaml.org> (accessed September 2010). Corsaro, OMG RFP mars/2008-05-03. <http://www.omg.org/cgi-bin/doc?mars/2008-05-03> (accessed September 2013) (restricted access).

[8] OMG, Extensible and Dynamic Topic Types for DDS. <http://portals.omg.org/ dds/sites/default/files/x-types_ptc_11-03-11.pdf> (accessed September 2013).

[9] D. Anicic, P. Fodor, S. Rudolph, N. Stojanovic, and EP-SPARQL: A unified language for event processing and stream reasoning, in: Proceedings of the 20th International Conference on World Wide Web, WWW '11, Hyderabad, India, ACM, New York, NY, USA, 2011, pp. 635–644.

[10] J. Aasman, Unification of geospatial reasoning, temporal logic, & social network analysis in event-based systems, in: Proceedings of the Second International Conference on Distributed Event-Based Systems (DEBS), 2008, pp. 139–145.

[11] S. Abdallah, Y. Raimond, the Event Ontology. <http:// www.motools.sourceforge.net/event/event.html> (accessed July 2013).

[12] A. Llaves, H. Michels, P. Maué, M. Roth, Semantic event processing in ENVISION, in: Proceedings of the 2nd International Conference on Web Intelligence, Mining and Semantics (WIMS), 2012, pp. 25:1–25:9.

[13] S. Sahoo et al., A Survey of Current Approaches for Mapping of Relational Databases to RDF. <http://www.w3.org/2005/Incubator/rdb2rdf/ RDB2RDF_SurveyReport_01082009.pdf> (accessed July 2013).

[14] P. T. Thuy, Y.-K. Lee, S. Lee, and XSD2RDFS and XML2RDF transformation: a semantic approach, in: Proceedings of the second International Conference on Emerging Databases (EDB).

[15] D. McGuinness, F. van Harmelen, OWL Web Ontology Language – Overview. <http://www.w3.org/TR/owl-features/> (accessed July 2013).

[16] F. Giunchiglia, P. Shvaiko, M. Yatskevich, and S-Match: an algorithm and an implementation of semantic matching, Semantic Web: Res. Appl. Lect. Notes Comput. Sci. 3053 (2004) 61–75.

[17] P. Bouquet, L. Serafini, S. Zanobini, Semantic coordination: a new approach and an application, in: The Semantic Web – ISWC 2003, Lecture Notes in Computer Science, vol. 2870, 2003, pp. 130–145.

[18] Gartner, Hype Cycles 2012. <http://www.gartner.com/technology/research/ hype-cycles/> (accessed July 2013).

[19] D. Anicic, P. Fodor, S. Rudolph, N. Stojanovic, and EP-SPARQL: A unified language for event processing and stream reasoning, in: Proceedings of the 20th International Conference on World Wide Web, WWW '11, Hyderabad, India, ACM, New York, NY, USA, 2011, pp. 635–644.

[20] D.J. Abadi, D. Carney, U. Çetintemel, M. Cherniack, C. Convey, S. Lee, M. Stonebraker, N. Tatbul, S. Zdonik, Aurora: A new model and architecture for data stream management, VLDB J. 12 (2) (2003) 120–139.