

Study of Distributed Data & Complex Event Processing System for Big Data

Neeraj Sharma^{1*} Dr. Mahaveer Sain²

¹ Research Scholar, Department of Computer, Science Maharishi Arvind University, Rajasthan

² Professor, Department of Computer Science & Informatics, Maharishi Arvind University, Rajasthan

Abstract – All through the long term, tremendous volumes of information are constantly created because of the expanding number of uses, proficient strategies are in this manner needed to decide the occasion examples of intrigue and oversee exceptionally unique occasions continuously. There has been expanding interest for dynamic frameworks inside Internet of Things, which can naturally respond to occasions that originate from different sources.

The Complex Event Processing (CEP) is an amazing innovation that can manage huge measure of information from different sources relying upon the consistency of information to create precise outcome to handle dynamic information continuously. In this manner, understanding existing CEP strategies and instruments is basic to build up a vigorous and viable CEP framework. In this paper, we had quickly depicted about occasion preparing, CEP with various motors and CEP for vulnerability. This paper surveyed CEP instruments accessible in the market from 2002 to 2019. It has been discovered that there are many popularized and open-source CEP devices in current market, where marketed devices are utilized for business knowledge reason and open-source instruments are generally utilized for scholastic purposes. The vast majority of the accessible preparing instruments are Query-based and not many are working with Machine learning in Complex Event Processing.

Keywords – Big data, Query language, Machine Learning, Complex event processing etc.

-----X-----

I. INTRODUCTION TO COMPLEX EVENT PROCESSING

In recent years the CEP has demonstrated to work as center device to handle constant information from sources to distinguish any conditions on solicitation and act immediately. CEP is additionally ready to handle huge erratic information, which is being created with the quickly developing programming applications or sensors from different sources. Such exponential information development causes the 5Vs in Big Data nature with outrageous speed of streaming information. It presents huge difficulties to beat issues progressively enormous information examination and dynamic. The Cisco determined that, by 2020, information will reach to 915 EB level (quintuple by 2020) from 171 EB in 2015. Though, information volume may reach out to 247 EB by 2020 from 25 EB in 2015. The information driven by the Internet of Things will arrive at 600ZB every year by 2020 from 145ZB every year in 2015.

1.1 Event Processing System

An Event is something important that happens or is relied upon to happen. An occasion happens

altogether or not in the least and is profound as it can influence specific activity. It is considered as stylish as it may be a situation to be genuine which might be advancement of an article in reality. For example, it very well may be a part for preparing business for exchanging, airplane flight plan, perused information from sensors, or it might be utilized to screen information of IT foundation, applications and middleware. In Figure 1 a significant level review of occasion preparing is appeared. An occasion, a noteworthy element that happens internal or external degree of business, which can create the solicitation of an office, the presentation of handling business data, and additionally the delivering of cutting edge report.

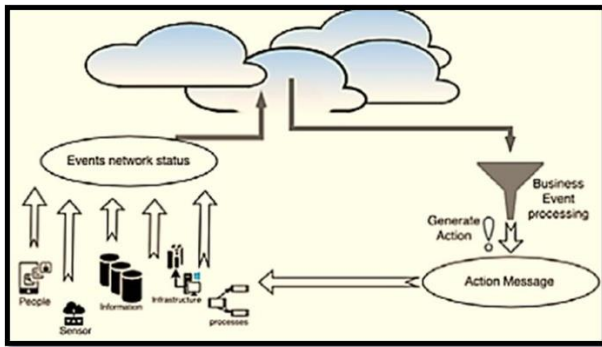


Figure 1. Showing Overview of Event Processing

The Event Processing manages at least one than one occasions to acquire objective to recognize the critical occasions inside the occasion cloud. Occasion handling capacities can be named into straightforward and complex with Event stream preparing commonly sorted by the occasions' multifaceted nature and modernity of preparing, as appeared in Figure 2.

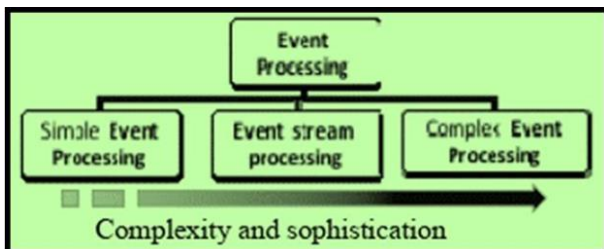


Figure 2. Showing Types of Event processing

2. Overview To CEP System

The Scalability, proficiency, idealness, vigor heterogeneity are the key attributes of CEP. It inputs unending, boundless occasion's stream from various sources to help ongoing information dealing with and recognizes a gigantic measure of occasions among low dormancy. Also, CEP frameworks are equipped for sifting, amassing, information to construe elevated level occasions and semantic material among occasions. Moreover, CEP can be utilized for wide scope of use, from basic observing to exceptionally complex applications, for example, extortion location and algorithmic exchanging.

3. Different CEP Engines

The various existing CEP methods with details information have been described to assist researchers for further improvement of CEP. The summary with restrictions of recent CEP engines for managing and monitoring complex events are presented in Table 1.

Table 1. Showing CEP Engines over Complex Events

Title	Application	Purpose	Restrictions
Runtime Verification of IoT Systems using Complex Event Processing	Internet of Things (IoT)	To detect failures at runtime by exploiting complex event patterns defined via predetermined event algebra.	undetected errors are missed due to network packet disorders
Cognitive Robot Programming using Procedural Parameters and Complex Event Processing	Robots	to improve the power and expressiveness of data processing systems in robots	highly focused on sensor data processing, relies on existing modules for performing actuation
Complex Event Processing for City Officers: A Filter and Pipe Visual Approach	Data processing pipelines in smart cities	A panel to let city workers to able to accurately design, create, test and implement their own set of data trough processing pipelines.	city officers require domain-specific knowledge
Knowledge-infused and consistent Complex Event Processing over realtime and persistent streams	Internet of Things (IoT) and Cyber-Physical Systems (CPS)	Accessible analytics over data stream and helps pan the velocity (real-time) and volume (persistent) dimensions.	Did not considered about uncertain events
Simulation of Real-time Vehicle Speed Violation Detection using Complex Event Processing	Traffic environments	Alert the infringe vehicles who are violating of speed limits	Did not consider vehicular multimodality, railway line traffic, bus transport options, pedestrian crossings and road block occurrences
Multi - stage monitoring of abnormal situation based on complex event processing	monitoring workspace	It processes high frequent data stream, tracking production process to identify production	Uncertainty has not been considered in state transition model
Efficient Context-Aware Nested Complex Event Processing over RFID Streams	RFID, sensors and mobile phones	To track and monitor anomalies and trends in event streams	Not experimented over some industry scenario and data sets.
Model4CEP	General	Situations detection in a user-friendly and intuitive way for a particular domain.Actions taken to notify interested users by social networking services or email.	Not able to model uncertainty in event patterns
Intelligent M2M	Automatic machine-to-machine communication	The incoming KPI rules are analyzed by the event processing rules in order to detect an event pattern that results in an 'Operating Event'	No CEP domain graphical definition, Event pattern graphical definition, No uncertainty supports, Elements represented as graphical nodes

4. Uncertain Data

The absence of precision and inadequate data on the data is known as weakness. The dynamic pattern of complex event dealing with faces various challenges because of free, missing, wrong, and equivocal information. These weaknesses rise because of nonappearance of data, dubiousness, precise errors, and insufficient information.

In programming building, this weakness makes sketchy data holds upheaval which causes it to diverge from the principal data. In the time of enormous data, weakness is the describing characteristics of data. Starting late, numerous cutting edge instruments have been set up to assemble and store gigantic volume of data, where in various occasions the data incorporates uncertain data with bumbles, commotion or may be deficient nuances. Moreover, a particular degree of weakness is shown by the persistent events, which perseveringly adulterates the accuracy of the introduction of employments.

Nowadays, various business affiliations, insightful foundations have gotten a handle on CEP and prescribed to control uncertainly without dismissing. There are various CEP applications,

for instance, blackmail acknowledgment, steady source gathering, framework's sensors, calculating trading, interference revelation, sound repeat conspicuous verification (RFID). In any case, the profitability and execution of these potential applications are affected by a couple of questionable data. There are authority working with these uncertainly, the diagram of relatively few CEP engines realized for uncertain events by the experts is presented in table 2. Title Application Purpose Limits Modeling and Formal Analysis of Probabilistic Complex Event Processing (CEP) Applications Business Process Management (BPM) to discover patterns within the events cloud providing the business manager by interesting information This algorithm needs to constraints so that it should be able to adopt the probabilistic timed automata (PTA) Predictive complex event processing based on evolving Bayesian networks Automatic online prediction To predict future states and take some actions in advance Low performance of the parameter updating and scalability issue of Bayesian networks Complex Event Processing under Uncertainty weather forecasting engine It uses BN to model uncertainty in rules and probability theory to manage uncertainty in events. To model a large number of uncertain rules it has scalability issue of Bayesian networks. Effective Privacy Preservation over Composite Events with Markov Correlations radio frequency identification (RFID) for privacy preservation Focuses on the privacy preservation of the composite events modeled by Markov chains due to the inherent uncertainty and correlations of data. Complex Event Processing on Uncertain Data Streams in Product Manufacturing Process Automatic production and manufacturing factories To monitor quality of products from group uncertain raw data frequently produced from the production or manufacture lines Modeling approach for situational event handling within production planning and control based on complex event processing Production, planning and control for carbon fiber reinforced plastic (CFRP) components. Agility, responsiveness, decision making and real-time reaction systems to handle complexity Not defined patterns parallel to the ongoing production processes

Automatic online prediction To predict future states and take some actions in advance Low performance of the parameter updating and scalability issue of Bayesian networks Complex Event Processing under Uncertainty weather forecasting engine It uses BN to model uncertainty in rules and probability theory to manage uncertainty in events. To model a large number of uncertain rules it has scalability issue of Bayesian networks. Effective Privacy Preservation over Composite Events with Markov Correlations radio frequency identification (RFID) for privacy preservation Focuses on the privacy preservation of the composite events modeled by Markov chains due to the inherent uncertainty and correlations of data.

Complex Event Processing on Uncertain Data Streams in Product Manufacturing Process Automatic production and manufacturing factories To monitor quality of products from group uncertain raw data frequently produced from the production or manufacture lines

Modeling approach for situational event handling within production planning and control based on complex event processing Production, planning and control for carbon fiber reinforced plastic (CFRP) components. Agility, responsiveness, decision making and real-time reaction systems to handle complexity Not defined patterns parallel to the ongoing production processes

Table 2. Showing CEP Engines for Handling Uncertain Events

Title	Application	Purpose	Limits
Modeling and Formal Analysis of Probabilistic Complex Event Processing (CEP) Applications	Business Process Management (BPM)	to discover patterns within the events cloud providing the business manager by interesting information	This algorithm needs to constraints so that it should be able to adopt the probabilistic timed automata (PTA)
Predictive complex event processing based on evolving Bayesian networks	Automatic online prediction	To predict future states and take some actions in advance	Low performance of the parameter updating and scalability issue of Bayesian networks
Complex Event Processing under Uncertainty	weather forecasting engine	It uses BN to model uncertainty in rules and probability theory to manage uncertainty in events.	To model a large number of uncertain rules it has scalability issue of Bayesian networks.
Effective Privacy Preservation over Composite Events with Markov Correlations	radio frequency identification (RFID) for privacy preservation	Focuses on the privacy preservation of the composite events modeled by Markov chains due to the inherent uncertainty and correlations of data.	
Complex Event Processing on Uncertain Data Streams in Product Manufacturing Process	Automatic production and manufacturing factories	To monitor quality of products from group uncertain raw data frequently produced from the production or manufacture lines	
Modeling approach for situational event handling within production planning and control based on complex event processing	Production, planning and control for carbon fiber reinforced plastic (CFRP) components.	Agility, responsiveness, decision making and real-time reaction systems to handle complexity	Not defined patterns parallel to the ongoing production processes

II. EXISTING VENDORS AND PRODUCTS OF CEP

Presently it has become enormous test because of gigantic measure of business exchanges for preparing streaming ongoing information in a programmed and composed strategy to create occasion driven framework. So as to perceive huge occasions from the continuous information, master framework is required. The CEP is a method created to deal with ongoing information and enormous information aggregate from various sources to get critical results or improvements from them. Barring a couple of insignificant limitations, the odds in this gear are a few. Nonetheless the innovation needs to perceive its ideal possible preparing for programmed business dependent on rules, inquiry or AI calculations and extra helpful blend for overseeing business information base.

It is immensely ridiculous to rely upon human to deal with, measure and imagine enormous information created from any business zone. Hence CEP instruments have uncovered its market noteworthiness for low inertness classification mix, affiliation and cycle contingent upon occasions assembled from continuous streaming information. Right now there are accessible apparatus for CEP is introduced for business and open sources. We have characterized the assessment angles for better perceivability to raise the significance of CEP instruments. We have demonstrated the synopsis of different free apparatuses in table 3 and business devices in table 4. Where, 'developer(s)' shows the individual or association anxious with parts of the device improvement. The field 'Name' referenced the name of the apparatuses. The field 'Preparing type' shows the buildup the CEP device which approach dependent on the framework was created.

Each apparatus is represented with various shade of square shape, where the left limits show when the device was delivered in market Yet, in a portion of the apparatuses the beginning date was anticipated in light of the fact that the precise proof was not introduced right now. The key/image 'obtained by' is utilized to demonstrate the left side instrument in square shape was gained by the organization of the correct side apparatuses. The appearance 'Follow on items' is utilized to show the devices accessible are needy to other device in market. (Open source) image is utilized to show the free accessible devices for buyers. At base time hub from 2002-2017 with keys in right side is drawn.

Table 3. Showing Summary of Existing Free CEP Tools

Developer	Name	Processing type
EsperTechInc	EsperTechNEsper	Query-based
Red hat	Drools Fusion	Inference Rule-based
WSO2	WSO2 Complex Event Processor	Query-based
Apache	Apache Flink	Query-based, Distributed Stream Computing Platform
Apache	Apache Samza	Neural-Network based
Striim	Striim	Neural-Network based
Apache	Apache Apex/ Data Torrent RTS	Query-based
EBay	EBay Pulsar	Query-based
Twitter	Twitter Heron	Query-based

Table 4. Showing Summary of Existing Commercial CEP Tools

Developer	Name	Processing type
Software AG	Apama Software AG	ECA Rule-based
TIBCO ("Tibco")	TIBCO LogLogic	Query-based
	Tibco Business Events	Inference Rule-based
	TIBCO Streambase	Query-based
IBM	WebSphere Business Events	ECA Rule-based
IBM	InfoSphere Streams	Query-based
Sybase - An SAP Company	SAP Sybase ASE	Query-based
Oracle	Oracle Complex Event Processing 10g	Query-based
FUJITSU	Software Interstage Big Data CEP Server	Query-based
Informatica	InformaticaRulePoint CEP	ECA Rule-based
One Market Data	OneTick CEP	Query-based
Microsoft	Microsoft StreamInsight	Query-based
Azure	Azure Stream Analytics	Query-based
Hitachi	Hitachi uCosminexus Stream Data Platform	Query-based
Vitria	Vitria Operational Intelligence	Query-based
Evam	Evam Event and Action Manager	Neural-Network based
Apache	Apache Strom	Query-based
SQLStream	SQLStream Blaze: s- Server	Query-based
SAS	SAS Event Stream Processing	Distributed Stream Computing Platform
Google	Google Cloud Dataflow	Query-based, Neural-Network
Apache	Apache Spark Streaming	Distributed Stream Computing Platform
Microsoft	Microsoft Trill	Query-based

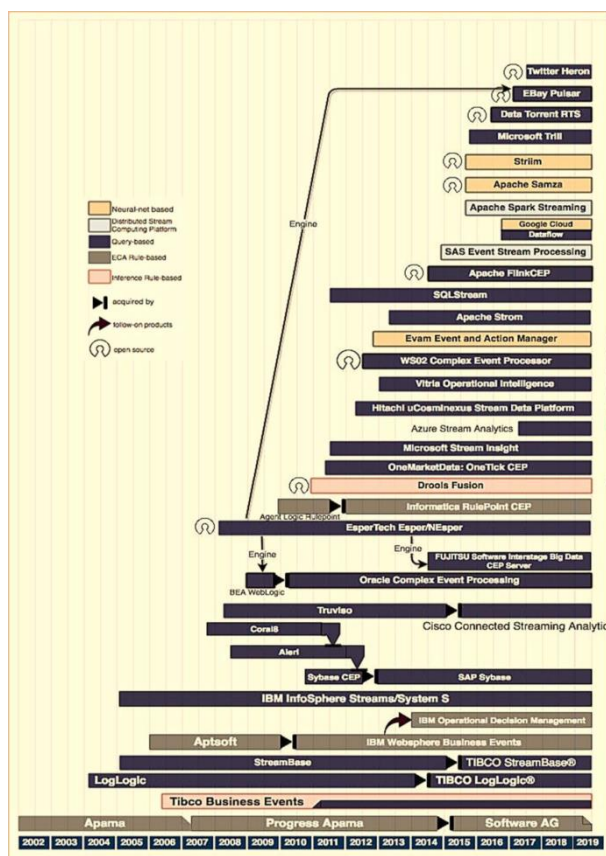


Figure 3. Showing Timeline of Existing Tools of CEP

3. CONVERSATION ON THE CHALLENGES OF CEP

The Many CEP motors and question dialects have been created to get unpredictable occasions. However, the current methodologies despite everything have restrictions to characterize and proliferate the vulnerability and impression from the majority of the constant applications. Change of ongoing information from crude information to valuable and reasonable data for end clients is very testing task for CEP. However, as of late, numerous CEP specialists proposed to deal with uncertainty. Be that as it may, considering to care for high vulnerability the majority of the current strategies runs in low adaptability and inconsequential execution.

It gets unsatisfactory for ongoing business frameworks for lack of ability of preparing the entering continuous occasions precisely. Uncertainly ordinarily happened because of hole between various information sources and genuine occasions frequencies. Subsequently, the current deterministic question dialects are not equipped for dealing with the unsure occasions effectively as inquiry language are restricted. It is imperative to conquer difficulties and give exact results from occasions to get assessment CEP market. The fundamental idea of occasion handling is understanding information, assessing that

information, determining an outcome and making a move dependent on that outcome. In view of handling ventures there are a few different reasons which make CEP a test as following:

3.1. Data Quality Assurance

The basic challenges to control the data quality of big data are scaling large amount of data on time, to obtain the best method to manage large scale of data and finally efficient techniques to clean those data. Most common data quality issue is due to unwanted data, missing data and data outliers.

I. Unwanted data and missing data

In this era of big data, it is fact that on generating large number of data and information may also generate large amount of unwanted data [50] and missing data [51]. Usually while cleaning data, if the cleaning data set is very small, it is easy to remove unwanted data and obtain useful information. And if the data set is very rough and big it is very hard to get the accurate results. On the one hand, it demands system to coordinate data in a short time; on the other hand, it also needs the method to build a rapid reaction to data in real time. Besides, the data may be filtered at a time node may become critical post processing data. Hence, it is quite challenging to understand and correlate among data sets and precisely control the worth and usefulness of data.

II. Data outliers

The information is made by at least one producing measures, which could either reflect action in the framework or perceptions gathered about substances. That is to say, anomalies are caused because of sensor information deviation from factors related and sound qualities. Exception location is a significant errand in Complex Event Processing for precision of information examination. It is trying to distinguish concealed exception, which can be organized known range, despite the fact that can't be illustrative for the influenced variable.

2. Evaluation of different kinds of data

It is one of the most crucial to deal with massive amount of heterogeneous information, which is beyond human being's capacity to digest. This type of information may contain different kinds of data and it is challenging to evaluate these data.

3. Lack of prior knowledge

Besides, overflow of semi-structured and unstructured data becomes challenging to construct its detained prescribed relations while analyzing the data. Furthermore, it is also problematic for these data required to be handled and monitored in real-time within satisfactory time to determine a priori

knowledge for upcoming streaming of the data in the structure of an endless streaming data.

4. Cope with resource constraints

The quantity of the data with quantity of measurement results needs to be processed, stored or transformed per time unit getting higher data rate. The CEP obtains input from different sources which is also big data and this big data amount may increase in future more gradually. Therefore, CEP tool should be able to overcome the challenge to process data and measurement results quantity.

5. Data privacy concerns

The Data protection issues are connected with the appearance of PCs in nearness. CEP apparatuses are utilized for preparing ongoing information with Data Privacy. The necessity for each undertaking is to conquered protection issues of successive irregularity with the intrigue to acquire however much data as could be expected. Protection concerns are basic to be esteemed to limit information assortment and data sharing. For example, in systems to distinguish squares it is fundamental to discover the reason for a disappointment and from the issue the sources should be isolated. It is enormously significant for arrange watchman to share information between organize administrators to follow the root of the assault. However, mystery and protection concerns made sharing of system information more troublesome and testing.

IV. CONCLUSIONS

In this period of Big information, specialists and engineers are focusing on the improvement of CEP to rise CEP business market position. 'Apama Software AG' device is still in market from 2010. During the most recent decade, designers delivered around too many CEP apparatuses in the market. These instruments are generally evolved utilizing rule-based inquiry to execute occasion progressions. This features the point 'Rules are all over the place'. Be that as it may, numerical formalisms can't immediate all the techniques for different occasion handling frameworks. Thus, we need AI calculations that can supplant specialists in producing rule designs.

A device, supporting a predictable and broadly perceived guideline language, could be created with more assets, since its market potential would be bigger than the one for an instrument supporting only one language in a divided standard scene.

There are a couple of difficulties looked by CEP designers, one of the most well-known issues being vulnerability, because of constant information event from various information sources. This information may bring grimy information (for

instance, Missing information or undesirable information) and information anomalies. There are likewise different difficulties, such as assessing various types of information with absence of earlier information, adapting to asset imperatives and information security concerns.

Nonetheless, CEP motor is usually the best answer for oversee and examine continuous occasions. CEP apparatuses are demonstrated to be the best arrangement in various spaces, for example, continuous business modules, media transmission, Capital Markets, sensor areas, Intelligence and Military, E-trade, Multiplayer Online Gaming zones and so on. Yet, right now, just a couple of apparatuses are accessible with AI capacities and there is wide extension for research for more CEP applications utilizing Machine Learning. The creators are working toward this path.

REFERENCES:

1. T. Lu, Z. Zha, and Z. Zhao (2017). "Multi-stage monitoring of abnormal situation based on complex event processing," *Procedia - Procedia Comput. Sci.*, vol. 98, no. September, pp. 1360-1368.
2. S. Peng and J. He (2017). "Efficient Context-Aware Nested Complex Event Processing over RFID Streams," *Web-Age Inf. Manag.*, vol. 9998, pp. 125-136.
3. J. Boubeta-puig, G. Ortiz, and I. Medina-bulo (2017). "Model4CEP: Graphical domain-specific modeling languages for CEP domains and event patterns," *Expert Syst. Appl.*, vol. 42, no. 21, pp. 8095-8110.
4. Drools, "Home page Drools," 2009. [Online]. Available: <https://www.drools.org/>.
5. H. D. B (2017). "Modeling and Formal Analysis of Probabilistic Complex Event Processing (CEP) Applications," *Eur. Conf. Model. Found. Appl.*, vol. 10376, pp. 248-263.
6. EsperTech, "EsperTech: Event Series Intelligence," 2006. [Online]. Available: <http://www.espertech.com/>.
7. Apache, "Home page Apache Samza," 2013. .
8. STRIIM, "Home Page," 2013. [Online]. Available: <http://www.striim.com/>. [Accessed: 20-Sep-2017].
9. Apache Apex TM (2017). "Home page Apache ApexTM." [Online]. Available: <https://apex.apache.org/>. [Accessed: 20-Sep-2017].
10. S. Murthy and T. Ng (2015). "Announcing Pulsar: Real-time Analytics at Scale," *Cloud, Data Infrastructure and Services, Open Source*, 2015. [Online]. Available: <http://www.ebaytechblog.com/2015/02/23/announcing-pulsar-real-time-analytics-at-scale/>.
11. S. Kulkarni, N. Bhagat, M. Fu, V. Kedigehalli, C. Kellogg, S. Mittal, J. M. Patel, K. Ramasamy, and S. Taneja (2015). "Twitter Heron: Stream Processing at Scale," in *SIGMOD'15*, pp. 239-250.
12. Software (2017). "Software AG Acquires Apama." [Online]. Available: <http://www.softwareag.com/corporate/company/apama.asp>. [Accessed: 18-Sep-2017].
13. "Tibco." [Online]. Available: <https://www.tibco.com/>.
14. IBM, "WebSphere Business Events Business: Event Processing Software." [Online]. Available: <https://www-1.ibm.com/software/integration/wbe/>.
15. Informatica, "Rule Point Complex Event Processing," 2010. [Online]. Available: https://www.informatica.com/products/data-integration/real-time-integration/rulepoint-complex-event-processing.html#fbid=aOGK9o_L9j9. [Accessed: 21-Sep-2017].
16. Microsoft, "developer.microsoft.com," 2009. [Online]. Available: <https://msdn.microsoft.com/en-us/>. [Accessed: 21-Sep-2017].
17. N. Mehdiyev, J. Krumeich, D. Enke, D. Werth, and P. Loos (2015). "Determination of Rule Patterns in Complex Event Processing Using Machine Learning Techniques," *Procedia Comput. Sci.*, vol. 61, pp. 395-401, 2015.
18. Z. Zheng, P. Wang, J. Liu, and S. Sun (2015). "Real-Time Big Data Processing Framework: Challenges and Solutions," *An Int. J. Appl. Math. Inf. Sci.*, no. 6, pp. 3169-3190.

Corresponding Author

Neeraj Sharma*

Research Scholar, Department of Computer, Science Maharishi Arvind University, Rajasthan

nesh787@rediffmail.com