ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

Efficient Query Processing in Geographic Web Search Engines

[1]Akanksha Singh, [2]Samarth Shekhar, [3]Bhagwati Ghbhane, [4]Prof. Latika Desai, [5]Prof. Santosh Chobe

[1]B. E. Student, Department of Computer Engineering, Dr. D. Y. Patil institute of Engineering and Technology, Pune, Maharashtra, India

[2]B. E. Student, Department of Computer Engineering, Dr. D. Y. Patil institute of Engineering and Technology, Pune, Maharashtra, India

[3]B. E. Student, Department of Computer Engineering, Dr. D. Y. Patil institute of Engineering and Technology, Pune, Maharashtra, India

[4]Assistant Professor, Department of Computer Engineering., Dr. D. Y. Patil institute of Engineering and Technology, Pune, Maharashtra, India

[5]Head of Department, Department of Computer Engineering., Dr. D. Y. Patil institute of Engineering and Technology, Pune, Maharashtra, India

ABSTRACT: Microblogging services became among the foremost modern services on world wide web at intervals the previous number of years. This LED to special increase in info size, speed, and applications. This paper presents Venus; a system that supports amount abstraction queries on microblogs. Venus supports its queries on a abstraction boundary R and a temporal boundary T, from that exclusively the top-k microblogs unit of measurement came back at intervals the question answer supported a spatio-temporal ranking operate. Supporting such queries desires Venus to digest many alternative amount microblogs in main-memory with high rates, yet, it provides low question responses and economical memoryutilization. to the current end, Venus employs: (1) a cost-effective in-memory spatio-temporal index that digests high rates of incoming microblogs in real time, (2) a ascendable question processor that prune the search space, R and T, effectively to supply low question latency on varied things in real time, and (3) a gaggle of memory improvement techniques that offer system administrators with utterly totally different selections legion|to avoid wasting} lots of vital memory resources whereas keeping the question accuracy nearly wonderful. Venus memory improvement techniques build use of the native arrival rates of microblogs to showing neatness shed microblogs that unit of measurement old enough to not contribute to any question answer. in addition, Venus can adaptively, in real time, alter its load shedding supported every the abstraction distribution and so the parameters of incoming question lots. All Venus components can accommodate utterly totally different abstraction and temporal ranking functions that unit of measurement able to capture the importance of each dimension otherwise wishing on the applications desires. thorough experimental results supported real Twitter info and actual locations of Bing search queries show that Venus supports high arrival rates of up to sixty four K microblogs/second and average question latency of 4 unit of your time..

KEYWORDS: Microblogs, spatial, location, temporal, performance, efficiency, scalability, memory optimization, social.

I. INTRODUCTION

Social media websites have grabbed large attention at intervals the last decade attributable to its growing quality and unprecedentedly big user base. The new wave of user-interactive microblogging services, e.g., tweets, comments on

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

Facebook or news websites, or Foursquare check-in's, has become the clear frontrunner at intervals the social media race with the foremost vital vary of users ever and highest users activity in consistent rates. for example, Twitter has 288+ Million active users World Health Organization generate 500+ Million daily tweets, whereas Facebook has one.35+ Billion users World Health Organization post 3.2+ Billion daily comments [3]. driven by the advances in wireless communication and so the standard of GPS-equipped mobile devices, microblogs service suppliers have enabled users to attach location knowledge with their posts. Thus, Facebook extra the alternatives of location check-ins and shut to where users can state a close-by location of their standing messages, Twitter automatically captures the GPS coordinates from mobile devices, per user permission, and Foursquare choices unit all around the location knowledge and so the whereabouts of its users. Consequently, a immoderateness of location knowledge is presently on the market in microblogs. we tend to tend to use of the provision of location knowledge in microblogs to support spatio-temporal search queries where users unit able to browse recent microblogs near their locations in real time. Users of our projected queries embody news agencies (e.g., CNN and Reuters) to possess a first-hand data on events throughout a positive house, advertising services to serve geo-targeted ads to their customers supported shut events, or individuals World Health Organization have to be compelled to grasp current activities throughout a positive house. for example, in New Style calendar month 2013, la Times reported [4] but people rush to Twitter for amount breaking news regarding Hub of the Universe Marathon explosions. Such users won't apprehend the suitable keyword or hash tag to travel searching for. Instead, they have to know the recently announce microblogs throughout a positive express house. Thus, our goal here is not to change the conventional keyword search in microblogs,but rather to provide another necessary search risk for localized microblogs. the answer of our spatio-temporal queries is also fed to various modules for any method, which may embody event detection, keyword search, entity resolution, sentiment analysis, or image.

II. RELATED WORK

1. "Efficient processing of window queries in the pyramid data structure,

Authors:W. G. Aref and H. Samet

Window operations operate the concept of style of queries which will be display throughout a abstraction data. samples of those window-based queries embrace the exist question (i.e., deciding whether or not or not or not a abstraction feature exists at intervals a window) and conjointly the report question, (i.e., news the identity of all the choices that exist at intervals a window). Algorithms area unit described for responsive window queries in &Ogr;(n log logT) time for a window of size n x n throughout a feature space (e.g., associate image) of size T x T (e.g., element elements). the importance of this result's that despite the actual fact that the window contains n2 element parts, the worst-case time quality of the algorithms is almost linearly proportional (and not quadratic) to the window diameter, and does not rely on different factors. the upper than quality bounds area unit achieved via the introduction of the unfinished pyramid system (a variant of the pyramid data structure) as a result of the underlying illustration to store abstraction choices and to answer queries on them.

2. "Mercury:A memory-constrained spatio-temporal real-time search on microblogs

Authors:A. Magdy, M. F. Mokbel, S. Elnikety, S. Nath, and Y.

This paper presents Mercury; a system for amount support of top-k spatio-temporal queries on microblogs, where users unit able to browse recent microblogs near their locations. With high arrival rates of microblogs, Mercury ensures amount question response at intervals a decent memory-constrained setting. Mercury bounds its search house to include only those microblogs that have got wind of intervals certain abstraction and temporal boundaries, throughout that only the top-k microblogs, in step with a spatio-temporal ranking operate, unit came at intervals the search results. Mercury employs: (a) a climbable dynamic in-memory index structure that is capable of digesting all incoming microblogs, (b) A question processor that exploits the in-memory index through spatio-temporal pruning techniques that deflate the amount of visited microblogs to return the final word answer, (c) To index size standardization module that dynamically finds and adjusts the minimum index size to substantiate that incoming queries area unit answered accurately, and (d) a load shedding technique that trades slight decrease in question accuracy for important storage savings.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

3. Logging infrastructure for data analytics at twitter,"
Authors:G. Lee, J. Lin, C. Liu, A. Lorek, and D. V. Ryaboy,
In recent years, there has been a substantial amount of labor on giant-scale data analytics victimization Hadoop-based platforms running on giant clusters of goods machines. A less-explored topic is but those data, dominated by application logs, unit of measurement collected and structured to begin with. throughout this paper, we have a tendency to tend to gift Twitter's production work infrastructure and its evolution from application-specific work to a unified "client events" log format, where messages unit of measurement captured in common, well-formatted, versatile Thrift messages. Since most analytics tasks take under consideration the user session as a result of the fundamental unit of study, we have a tendency to tend to pre-materialize "session sequences", that unit of measurement compact summaries that will associateswer an outsize class of common queries quickly. the event of this infrastructure has economical log assortment and knowledge analysis, thereby rising our ability to quickly experiment and retell on varied aspects of the service.

4. Large-scale machine learning at twitter
Authors:J. Lin and A. Kolcz,
The success of data-driven solutions to difficult problems, at the aspect of the dropping costs of storing and method immense amounts of knowledge, has junction rectifier to growing interest in large-scale machine learning. This paper presents a case study of Twitter's integration of machine learning tools into its existing Hadoop-based, Pig-centric analytics platform. we have a tendency to begin with Associate in Nursing outline of this platform, that handles "traditional" data deposit and business intelligence tasks for the organization. The core of this work lies in recent Pig extensions to provide prognostic analytics capabilities that incorporate machine learning, targeted specifically on supervised classification. specifically, we have acknowledged random gradient descent techniques for on-line learning and ensemble ways as being extraordinarily amenable to scaling bent huge amounts of knowledge. In our deployed resolution, common machine learning tasks like data sampling, feature generation, training, and testing are going to be accomplished directly in Pig, via painstakingly crafted loaders, storage functions, and user-defined functions.

5. "Earlybird: Real-time search at twitter,
Authors:M. Busch, K. Gade, B. Larson, P. Lok, S. Luckenbill, and J. Lin,
The web currently is a lot of and a lot of characterized by social and amount signals, that we tend to tend to believe represent two frontiers in information retrieval. throughout this paper, we tend to tend to gift Early bird, the core retrieval engine that powers Twitter's amount search service. tho' Early bird builds and maintains inverted indexes like nearly all fashionable retrieval engines, its index structures dissent from those designed to support ancient net search. we tend to tend to explain these variations and gift the reason behind our vogue. A key demand of amount search is that the power to ingest content quickly and build it searchable directly, whereas at identical time supporting low-latency, high-throughput question analysis. These demands unit of measurement met with a single-writer, multiple-reader concurrency model and additionally the targeted use of memory barriers. Early bird represents some extent among the design space of amount search engines that has worked well for Twitter's desires. By sharing our experiences, we tend to tend to hope to spur additional interest and innovation throughout this exciting space.

III. PROPOSED SYSTEM

We propose effective memory improvement techniques: (1) we tend to tend to associate anaalytically develop an index size standardisation technique that achieves vital memory savings (up to 50%) whereas not sacrificing the question answer quality (more than 99% accuracy). the foremost set up is to use the vary of arrival rates per regions.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

The planned system contains following process:

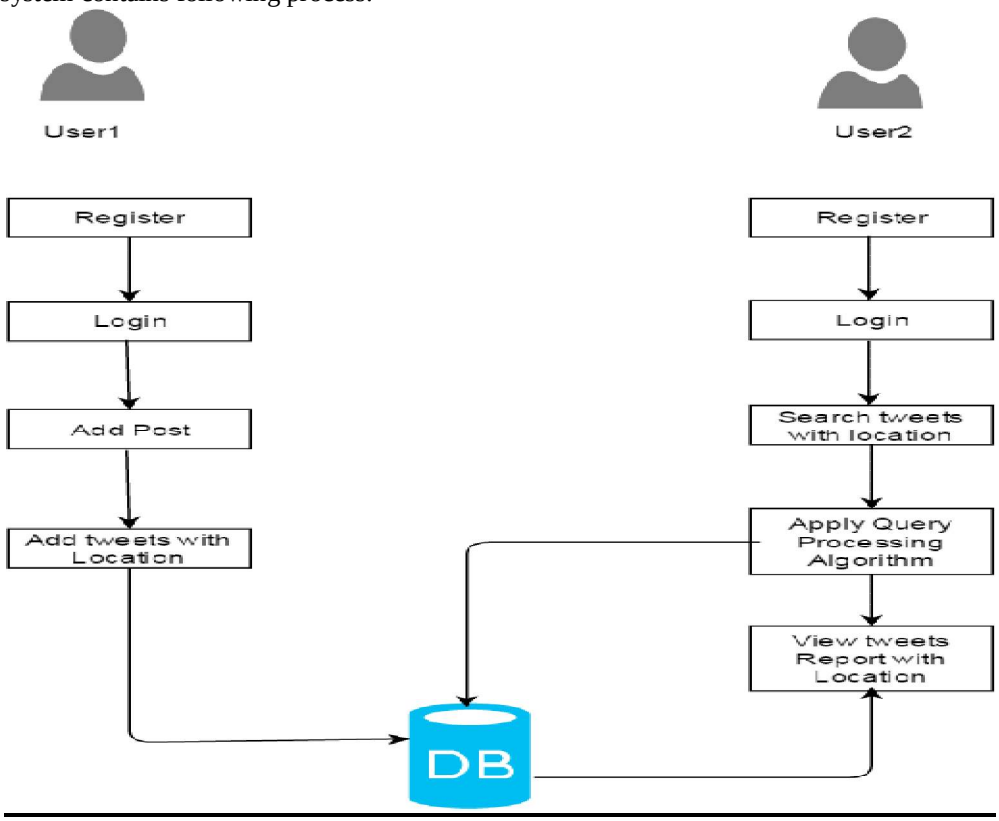


Figure: Proposed System Architecture

ADVANTAGES OF PLANNED SYSTEM:

1. Venus can use fully totally different ranking functions to be able to serve wants of varied applications.
2. These techniques catch the abstraction distribution of the incoming queries to boot as a result of the abstraction access patterns of the hold on microblogs so as that they produce the storage overhead to its least levels (up to 80% less storage) whereas allow to answer queries with nearly sensible accuracy (more than 99 % in all cases).

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

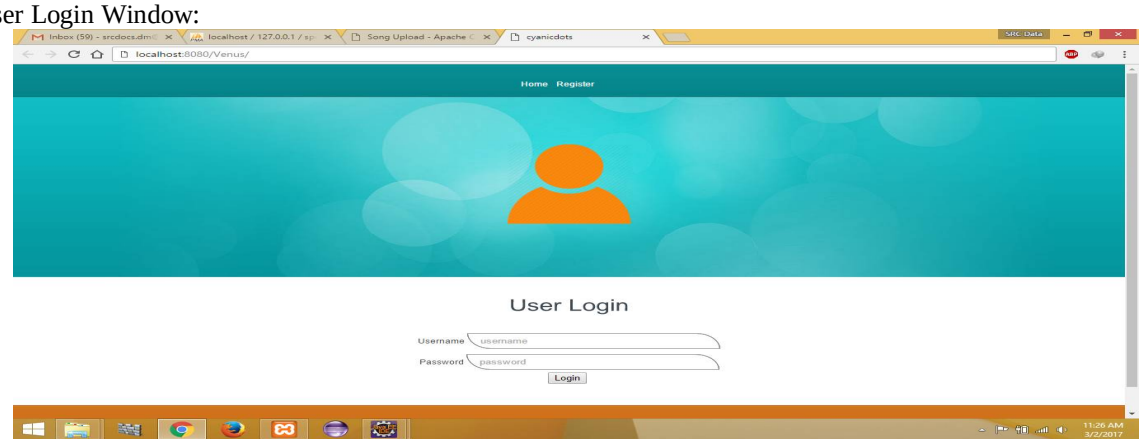
(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

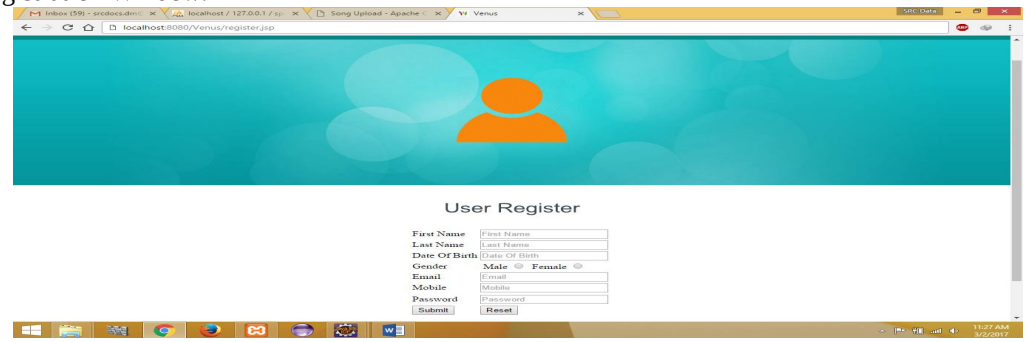
Vol. 5, Issue 4, April 2017

IV. RESULT ANS DISCUSSION

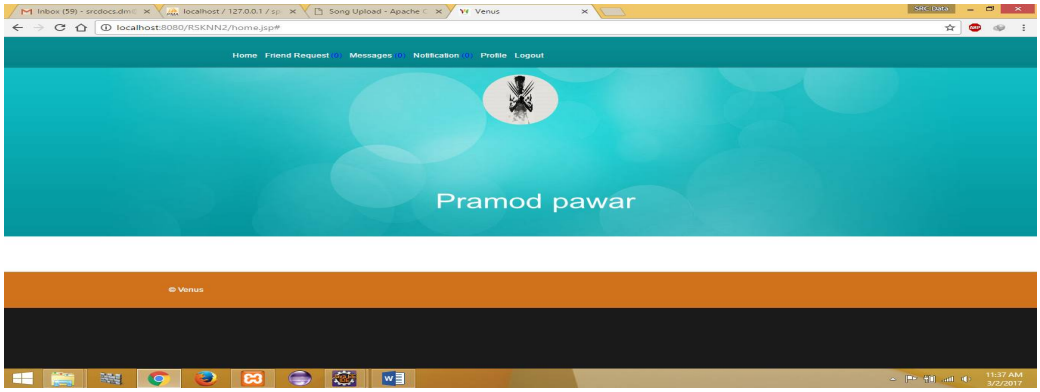
1. User Login Window:



2. User Registration Window:



3. User Home Window:



ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

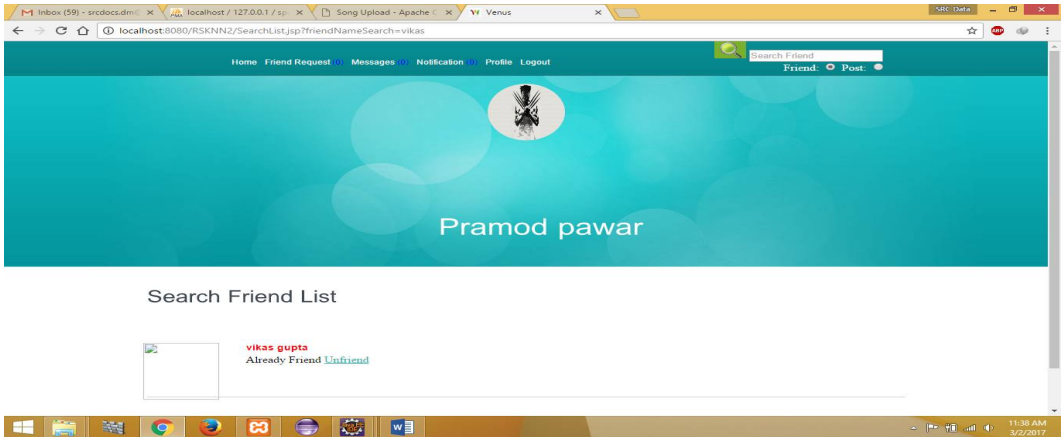
International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

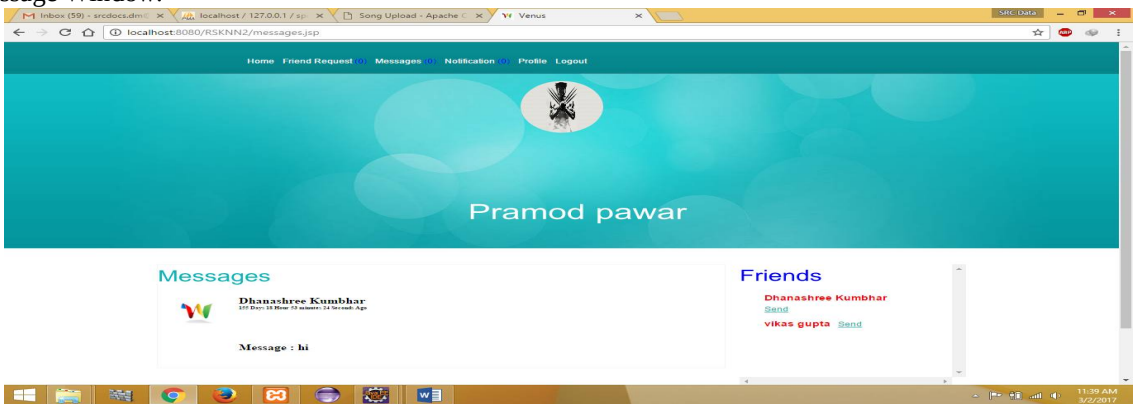
Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

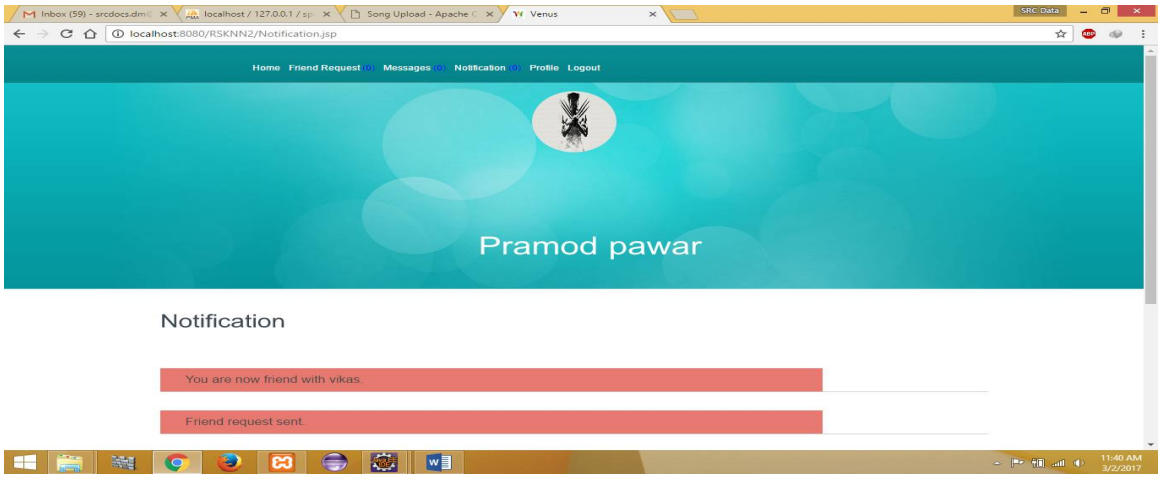
4. Friend Search Window:



5. Message Window:



6. Notification Window:



ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

V. CONCLUSION

We have presented Venus; a system for fundamental quantity support of spatio-temporal queries on microblogs, where users request a gaggle of recent k microblogs near their locations. Venus works beneath a troublesome setting, where microblogs arrive with really high arrival rates. Venus employs economical in-memory compartmentalization to support up to 64 Kmicroblogs/second and spatio-temporal pruning techniques
to provide fundamental quantity question response of 4 unit of time. In addition, effective load shedding modules are used to well shed the useless data whereas providing just about wonderful question accuracy.

REFERENCES

[1] (2013). Twitter Statistics [Online]. Available: http://expandedramblings. com/index.php/march-2013-by-the-numbers-a-few-amazing-twitter-stats/
[2] (2014). Twitter Data Grants, 2014 [Online]. Available: https://blog. twitter.com/2014/introducing-twitter-data-grants
[3] (2012). Facebook Statistics [Online]. Available: https://www.facebook. com/business/power-of-advertising
[4] (2013). After Boston Explosions, People Rush to Twitter for Breaking News [Online]. Available: http://www.latimes.com/business/ technology/la-fi-tn-after-boston-explosions-people-rush-to-twitter- for-breaking-news-20130415,0,3729783.story
[5] W. G. Aref and H. Samet, "Efficient processing of window queries in the pyramid data structure," in Proc. 9th ACM SIGACT-SIGMOD-SIGART Symp. Principles Database Syst., 1990, pp. 265–272.
[6] A. Magdy, M. F. Mokbel, S. Elnikety, S. Nath, and Y. He, "Mercury: A memory-constrained spatio-temporal real-time search on microblogs," in Proc. IEEE 30th Int. Conf. Data Eng., 2014, pp. 172–183.
[7] G. Lee, J. Lin, C. Liu, A. Lorek, and D. V. Ryaboy, "The unified logging infrastructure for data analytics at twitter," Proc. Very Large Data Base, vol. 5, no. 12, pp. 1771–1780, 2012.
[8] J. Lin and A. Kolcz, "Large-scale machine learning at twitter," in Proc. ACM SIGMOD Int. Conf. Manage. Data, 2012, pp. 793–804.
[9] M. Busch, K. Gade, B. Larson, P. Lok, S. Luckenbill, and J. Lin, "Earlybird: Real-time search at twitter," in Proc. IEEE 28th Int. Conf. Data Eng., 2012, pp. 1360–1369.
[10] C. Chen, F. Li, B. C. Ooi, and S. Wu, "TI: An efficient indexing mechanism for real-time search on tweets," in Proc. ACM SIGMOD Int. Conf. Manage. Data, 2011, pp. 649–660.
[11] Ankit Lodha, Clinical Analytics – Transforming Clinical Development through Big Data, Vol-2, Issue-10, 2016
[12] Ankit Lodha, Agile: Open Innovation to Revolutionize Pharmaceutical Strategy, Vol-2, Issue-12, 2016
[13] Ankit Lodha, Analytics: An Intelligent Approach in Clinical Trail Management, Volume 6 ,Issue 5 , 1000e124