An Introduction to Real Time Processing and Streaming of Wireless Network Data

MohitMaske[1], Dr. Prakash Prasad[2]

M.E. Student, Department of Computer Technology, Priyadarshini College of Engineering, Nagpur (MH), India [1]

HOD, Department of Computer Technology, Priyadarshini College of Engineering, Nagpur (MH), India [2]

Abstract: Growth of the internet, Lead to Cloud Computing, Mobile Network and Internet of Things increasing rapidly, big data is becoming a hot-spot in recent years. Data generated from many sources made huge demand for storing, managing, processing and querying on various stream. This gives rise to data processing in our daily life such as mobile devices, RFID and Wireless sensors, which aims at dealing with billions of users interactive data. At the same time real time processing is intently needed in integrated system. An entire system is built on one such application Storm, associated with Sql Stream. To ensure the practical applicability and high efficiency, A Simulation System is established and shown acceptable performance in various expressions using data sheet. It proved that data analysis system for stream and real time processing based on storm can be used in various computing environment.

Keywords: Dataset, Real time processing, Distributed Computing, Sql Streaming.

I.   INTRODUCTION

We are aiming at an entire system including data generator, data processing and data storage based on storm, which is novel distribution real time computing system. Real time processing is a fast and prompt data processing technology that combines data capturing, data processing and data exportation together [1].

The main purpose of real time data processing and stream is to realize an entire system that can process mesh data in short time [4]. The truth is the value of the Internet of things only comes from the astounding mass of data it's bound to produce, and there will be no money made until the data-processing market is ready to take on the challenge [3].

It has been widely used in variety of fields such as Big Science, RFID, social networks, and internet search indexing astronomy and so on. The Social network Facebook and twitter serve billions of page views per day and have billions of article and photos to store. The traditional application of data processing is distributed system [1].

Stream processing is a technology that allows for the collection, integration, analysis, visualization, and system integration of data, all in real time, as the data is being produced, and without disrupting the activity of existing sources, storage, and enterprise systems [3].

The goal of real-time processing is to provide solutions that can process Sql Stream data very fast and interactively. Stream processing deals with problems that their input data must be processed without being totally stored.

Data stream are now very common. Log Streams, Click steams, Message streams, and Event streams are some good examples but the standard of data analysis is completely focus on the data generator and data storage and processing[4].Many believe that, just as database management needed Oracle and big data needed Hadoop, the Internet of things (or the *really* big data) needs stream processing. The framework process the input data and the output results are available only when all of the computation in done.

The Advantage of this system is high fault tolerance, low-Cost, High reliability, High extensibility and high efficiency.

II.   ISSUES WITH STREAMING

Issues with streaming not perform real time data processing for generate, manage and produce data. This storm stream processing is used for frameworks like Twitter social networking sites. It has been widely used in variety of fields such as Big Science, RFID, social networks, and internet search indexing astronomy and so on [1]. The Social network Facebook and twitter serve billions of page views per day and have billions of article and photos to store. The traditional application of data processing is distributed system. On this organized real time dataprocessing system produce data on dataset and other tools and [1] according to simulation experiment, the system can be easily and widely used in many territories such as cloud computing and data mining.

Handling the Continuous flow of data streams:-

This is a data management issue. Traditional database management systems are not capable of dealing with continuous high data rate. Novel indexing storage and querying technique are required to handle this non stopping fluctuated flow of information streams [9].

Unbounded memory requirements due to the continuous flow of data streams: -

Machine learning techniques represent the main source of data mining algorithms. Most of machine learning methods

require data to be resident in memory while executing the analysis. Due to the huge amounts of the generated streams, it is absolutely a very important concern to deign space efficient techniques that can have only one look or less over the incoming stream [9].

Second thing issues with streaming, Hadoop is not used for these non-batch workloads, new solutions are proposed to these new challenges. Although MapReduce does not support stream processing, but MapReduce can partially handle streams using a technique known as micro-batching [4]. The idea is to treat the stream as a sequence of small batch chunks of data. So it all drawback support suitable for storm application. Storm integrates with the queuing and database technologies already use.

III.   LITERATURE REVIEW

Purpose of real-time data processing and stream is to realize an entire system that can process mesh data in short time [1].

Optionally use Data mining is the process of automatically extracting useful information and relationships from immense quantities of data [3]. an entire system including data generator, data processing and data storage, data capturing based on storm, which is novel distribution real-time computing system.

Our survey on this topic to data stream and click stream event data is one the most important sources of information in website usage. In survey SQL stream is most important process use in this related topic. The core part of the real time data processing system is our application storm. Storm is a free and open source distributed real time computation system. Storm has many use cases: real time analytics, online machine learning, continuous computation, distributed RPC, ETL,   and more.

We are Search Some Data Stream like Event Stream, Logs Stream, Message Stream and Click Stream. Event stream processing, or ESP, is a set of technologies designed to assist the construction of event-driven information systems [2]. ESP technologies include event visualization, event databases, event-driven middleware, and event processing languages, or complex event processing (CEP). In practice, the terms ESP and CEP are often used interchangeably. ESP deals with the task of processing streams of event data with the goal of identifying the meaningful pattern within those streams, employing techniques such as detection of relationships between multiple events, event correlation, event hierarchies, and other aspects such as causality, membership and timing [3].

Data stream management system (DSMS) is a computer program to manage continuous data streams. It is similar to a database management system (DBMS), which is, however, designed for static data in conventional databases. A DSMS also offers a flexible query processing so that the information need can be expressed using queries. However, in contrast to a DBMS, a DSMS executes a continuous query that is not only performed once, but is permanently installed. Therefore, the query is

continuously executed until it is explicitly uninstalled. Since most DSMS are data-driven, a continuous query produces new results as long as new data arrive at the system.

The formulation of queries is mostly done using declarative languages like SQL in DBMS. Since there are no standardized query languages to express continuous queries, there are a lot of languages and variations. However, most of them are oriented to SQL like the Continuous Query Language (CQL), StreamSQL or EPL.
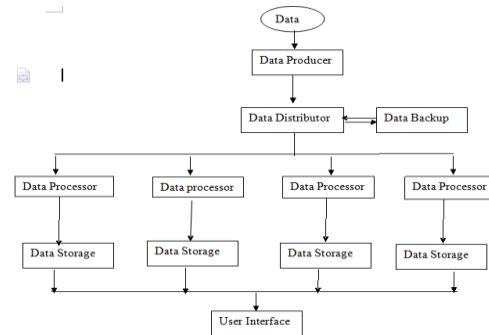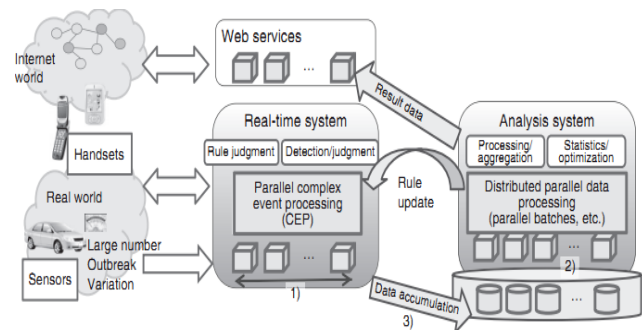

Fig: - General constitution of System

Figure 1
Overall picture of massive data processing on cloud.

IV.   RESEARCH METHODOLOGY

Since there is only one server under computation such as: IIS (Internet Information Server). It is for windows server is a flexible, secure and manageable web server hosting. The result show system can perform well under such environment.

In this research system performs several methodologies to be used as: - Sql Stream, Internet of Things, and Data mining and Real Time Processing.

SQL (Sql Stream):-

SQLstream s-Server remains the only SQL standards platform for data stream management. There's a reason for that. It's much simpler to build a proprietary platform and languages without the constraints of standards compliance. However, ultimately, this is a Fault Tolerance. Longevity and enterprise-class adoption require standards, stability, reliability, and importantly, a platform that can automate

the hard stuff such as distributed query optimization. SQL doesn't imply structured data; it's equally applicable for processing unstructured data. It's just the underlying data management technology that differs [3].

Stream Processing:-

Stream processing is a computer programming paradigm, related to SIMD (single instruction, multiple data), that allows some applications to more easily exploit a limited form of parallel processing. Such applications can use multiple computational units, such as the FPUs on a GPU or field programmable gate arrays (FPGAs),[1] without explicitly managing allocation, synchronization, or communication among those units.

The stream processing paradigm simplifies parallel software and hardware by restricting the parallel computation that can be performed. Given a set of data (a stream), a series of operations (kernel functions) is applied to each element in the stream. Uniform streaming, where one kernel function is applied to all elements in the stream, is typical. Kernel functions are usually pipelined, and local on-chip memory is reused to minimize external memory bandwidth. Since the kernel and stream abstractions expose data dependencies, compiler tools can fully automate and optimize on-chip management tasks. Stream processing hardware can use score boarding, for example, to launch DMAs at runtime, when dependencies become known.

Real-time:-

Stream processing when all is said and done is just a technology, an enabler for other things. What matters to Enterprises is the ability to deliver business value faster, and to improve operational efficiency through real-time actionable intelligence and operational process automation. Businesses are increasingly concerned about their ability to deliver real-time systems, whether for competitive differentiation, operational efficiency or optimal customer experience. The technology is not the major issue, it's what it offers to customers that really matters. Stream processing offers real-time systems and automated processes with the lowest cost of performance [2].

Internet of Things:-

The combination of Data Mining and the Internet is moving us into a new age of interaction with and understanding of the physical world around us. With the promise of real-time apps for connected cars, home automation, personal health, the industrial internet and smarter cities, we're all intrigued as to what the Internet of Things offers. Real-time is a key requirement. For example, the ability to respond to a sensor event in real-time, to deliver connected car infrastructure that offers real-time updates, but importantly, the ability to automate processes in real-time. This is where a data stream management platform comes into play. There are of course

many other components required, but the Internet of Things is being built on the full panoply of Big Data, Internet and Cloud technologies [3].

Data Mining:-

Data mining is process of automatically extracting useful information and relationships from immense quantities of data.

V.   CONCLUSION

In this survey we research many things of related data streaming and SQL stream and data stream, log stream and some various kind dataset stream.We demonstrated to issues about data streaming.The core module that represents the interface to the database has a number of important features that were critical to the successful operation of the system. No pre-processing is needed. The raw flow and occupancy measurements are used directly and are assumed.

This is can be widely used many territories such as cloud computing, data mining, use sql stream process and so on.

REFERENCES

[1] Wenjie Yang,Xingang Liu and Lan Zhang,"Big Data Real Time Processing Based on Storm"2013 12th IEEE International Conference on Trust, security and privacy.

[2] "Storm Homepage." [Online]. Available: http://storm-project.net. [Accessed: 01-Dec-2013].

[3] "Sql Stream Process:" [Online]. Available: "Stream Processing with SQL SQLstream Big Data Stream Processing & Operational Intelligence for the Internet of Things" http://SqlStream.com/stream processing" [Accessed: 01-Dec-2014].

[4] Saeed Shahrivari, and Saeed Jalili, Beyond Batch Processing: Towards Real-Time and Streaming Big Data

[5] H. Yang and S. Fong, "Incremental optimization mechanism for constructing a decision tree in data stream mining," Mathematical Problems in Engineering, vol. 2013, Article ID 580397, 14 pages, 2013. View at Publisher · View at Google Scholar.

[6] T. White, Hadoop: The definitive guide. Yahoo Press, 2012.

[7] M. A. Beyer and D. Laney, "The Importance of 'Big Data': A Definition," Jun. 2012.

[8] D. Dai, X, Li, C. Wang, M. Sun, X. Zhou, "Sedna: A Memory Based Key-Value Storage System for Real-time Processing in Cloud," 2012 IEEE International Conference on Cluster Computing Workshops, pp. 48-56, Sept. 2012.

[9] M. Zaharia, M. Chowdhury, T. Das, A. Dave, J. Ma, M. McCauley, M. J. Franklin, S. Shenker, and I. Stoica, "Resilient distributed datasets: a fault-tolerant abstraction for in-memory cluster computing," in Proceedings of the 9th USENIX conference on Networked Systems Design and Implementation, 2012.

[10] Mohamed Medhat Gaber, Arkady Zaslavsky and Shonali Krishnaswamy, "Mining Data Stream: A Review" Complete reference about streaming issues.

[11] H. Yang and S. Fong, "Incremental optimization mechanism for constructing a decision tree in data stream mining," Mathematical Problems in Engineering, vol. 2013, Article ID 580397, 14 pages, 2013. View at Publisher · View at Google Scholar.

BIOGRAPHIES

Mohit M. Maske is a M.E. Student of Priyadarshini College of Engineering, Nagpur (MH). His completed his Bachelor of Engineering in Shree Shankarprasad Agnihotri College of Engineering, Wardha. His got aggregate with 67% in whole year of B.E. and also completed of Software Development Training Course in Hyderabad Institute. Now Study of Master Degree from This College in Branch of Wireless Communication and Computing. And this paper all survey with Dr. Prakash Prasad.

Dr. Prakash S. Prasad obtained Ph.D. degree in Computer Science & Engineering. He has published more than 16 papers in National and International Conferences. He is Member of IEEE, ISTE and IACSIT. He has completed his bachelor's degree in 1997 and Master's Degree in 2007.He is currently working as Associate Professor at Priyadarshini college of Engineering and Head of The Department of Computer Technology. He is having 16 Years of teaching experience and his interests include network security, Operating System and System Software.