Scientific Journal of Impact Factor (SJIF): 4.72

e-ISSN (O): 2348-4470
p-ISSN (P): 2348-6406

International Journal of Advance Engineering and Research Development

Volume 4, Issue 6, June -2017

WEATHER SENSING DATA RECOGNIZATION USING HADOOP FRAMEWORK

Poonam Shinde[1], Rupali Mhase[2], Sneha Pawar[3], Nayan Soudagar[4]
Prof. Shubhangi vairagar[5]

[1,2,3,4,5] *Computer Engineering, Siddhant College of Engineering, Pune*

Abstract — *The assets of remote senses digital world daily generate Big volume of period of time information (mainly remarked the term "Big Data"), wherever insight data incorporates a potential significance if collected and mass effectively. In today's era, there's an excellent deal additional to period of time remote sensing Big information than it looks initially, Associate in Nursingd extracting the helpful data in an economical manner leads a system toward a significant procedure challenges, like to investigate, aggregate, and store, wherever information area unit remotely collected. Keeping visible the on top of mentioned factors, there's a desire for planning a system design that welcomes each real-time, yet as offline processing. Therefore, during this paper, we tend to propose period of time Big information analytical design for remote sensing satellite application. The planned design contains 3 main units, like 1) remote sensing Big information acquisition unit (RSDU); 2) processing unit (DPU); and 3) information analysis call unit (DADU). First, RSDU acquires information from the satellite and sends this information to the bottom Station, wherever initial process takes place. Second, DPU plays an important role in design for economical process of period of time Big information by providing filtration, load leveling, and multiprocessing. Third, DADU is that the higher layer unit of the planned design, that is answerable for compilation, storage of the results, and generation of call supported the results received from DPU. The planned design has the potential of dividing, load leveling, and multiprocessing of solely helpful information. Thus, it leads to expeditiously analyzing period of time remote sensing Big information exploitation earth observatory system. What is more, the planned design has the potential of storing incoming information to perform offline analysis on for the most part hold on dumps, once needed. Finally, an in depth analysis of remotely detected earth observatory Big information for land and ocean space area utit provided exploitation Hadoop and ocean space, additionally, varied algorithms area unit planned for every level of RSDU, DPU, Associate in Nursingd DADU to find land yet as ocean space to elaborate the operating of an design.*

Keywords*- Big Data, data analysis decision unit (DADU), data processing unit (DPU), land and sea area, offline, real-time.*

I. INTRODUCTION

RECENTLY, a great deal of interest in the field of Big Data and its analysis has risen, mainly driven from extensive number of research challenges strappingly related to bonfire applications, such as modeling, processing, querying, mining, and distributing large-scale repositories. The term "Big Data" classifies specific kinds of data sets comprising formless data, which dwell in data layer of technical computing applications and the Web. The data stored in the underlying layer of all these technical computing application scenarios have some precise individualities in common, such as 1) large-scale data, which refers to the size and the data warehouse; 2) scalability issues, which refer to the application's likely to be running on large scale (e.g., Big Data); 3) sustain extraction transformation loading (ETL) method from low, raw data to well thought-out data up to certain extent; and 4) development of uncomplicated interpretable analytical over Big Data warehouses with a view to deliver an intelligent and momentous knowledge for them . Big Data are usually generated by online transaction, video/audio, email, number of clicks, logs, posts, social network data, scientific data, remote access sensory data, mobile phones, and their applications . These data are accumulated in databases that grow extraordinarily and become complicated to confine, form, store, manage, share, process, analyze, and visualize via typical database software tools

II. RELATED WORK

This section provides the detail summary of the previous work done in the remote sensing real time big data. The digital world generating the high amount of the data continuously, current technology and the tools to store and analyze the large amount of data not an easy task, since it is not able to extract the needed data sets. So there is a need of an architecture that can analyze both the offline data as well as real time data sets. There is an influential benefit in the business enterprise by obtaining the required information from the Big data than sample data sets. Some of the areas that are described below where big data can play very important role. Understanding the earth atmosphere or environment

requires large volume of information or data gathered from different sources, such as air and water quality monitoring sensors, amount of oxygen, co2 and the other gases present in the air, remote access satellite for the observing the characteristics of the earth and so on. In the healthcare scenarios, there is large amount of the data about the medications, patients, medical history and other details gathered by the medical practitioner. The above mentioned data is very complex in nature, there is a chances of missing the important data. Day by day the data becoming very large by social networking, online streaming, system logs, mails and remote data, it will be very difficult to compute massive amount of data. Main problem is how to store the large amount of data i.e. big data and what data is to keep and what data is to be discarded, extracting the useful data from the big data is the challenging task. Most of the data is generated by the streaming data. In data stream model, the data will arrive at a very high speed and the algorithm have to process them. This data stream causes several challenges in design of the data mining algorithms. First, algorithm has to make use of less number of resources. Second, it can deal with data that can change over time. Resources are managed in an efficient and low cost way, by the green computing. Green computing is the process or study to use the computing resources in an efficient way. Here, the problem is not only the scaling issue but also error handing, lack of structure, heterogeneity, privacy, visualization and timeliness. The challenge is to design a high performance computing systems that can be able integrate resources from different location. Even though the cloud computing systems shown high level performance for RS applications, there are challenges still remaining regarding energy and the time consumption. The big challenge emerges when collecting and the managing Remote Sensing (RS) big data. The RS data are collected form spacecraft, airplanes, satellite and other sensing devices. Remote sensing data growing explosively, we have entered in the period of very high resolution, observation of the earth. Remote sensing data also considered as a "Big Data". With the advance sensors we can take even high spatial resolution images, spectral resolution and also temporal resolution. The advancement in the technology of the computers and the remote sensing devices increases a massive growth remote sensing data. The earth observatory data that is streamed from the spacecraft approximately around 1.7GB, this data is collected by single satellite and increased many terabytes per day. The global records of observatory data of the earth would exceed to one Exabyte, according to the OGC statistics. Various standard format data sets of remote sensing are stored in structured files, the formats including ASCII, HDF,netCDF and so on. Different organization have different standard format of the data sets, different format of data has its own format libraries and operation interfaces. Huge amount of data need to compute in an efficient way and only the useful information need to be extracted from the big data. So there was a need of the architecture for filtering the data, load balancing, aggregating and the decision analysis.

III. Data Model and Description

Data Description

Data in this application possesses information about the graphical objects regarding geographic regions which are in binary format.The offline data which is supposed to be in random formats such as numerical,textual or graphical binary.We are accessing data from sensors like primary camera,infrared as well as SONAR,RADAR,accelerometer,magnetometer,consideration of the each data is done in same way.

Data objects and Relationships

Data objects mainly RSDU,DPU and DADU.Firstly, RSDU gathers data from the satellite and transmit this data to the Base Station, where initial processing takes place. Later DPU plays an essential role in architecture for efficient processing of real-time Big Data by providing filtration, load balancing, and parallel processing. Then DADU is responsible for compilation, storage of the results, and generation of decision based on the results received from DPU.

Description of function :

First,Row Data acquisition collect data from the satellite and sends this data to the Base Station, where initial processing takes place.Filtration identifies the useful data for analysis since it only allows useful information, whereas the rest of the data are blocked and are discarded.using load balancing dividing the whole filtered data into parts and assign them to various processing servers.DADU is the which is responsible for compilation,

IV. SYSTEM ARCHITECTURE

A remote sensing Big Data analytical architecture, which is used to analyze real time, as well as offline data. At first, the data are remotely preprocessed, which is then readable by the machines. Afterward, this useful information is transmitted to the Earth Base Station for further data processing.

Earth Base Station performs two types of processing, such as processing of real-time and offline data. In case of the offline data, the data are transmitted to offline data-storage device.

The incorporation of offline data-storage device helps in later usage of the data, whereas the real-time data is directly transmitted to the filtration and load balancer server, where filtration algorithm is employed, which extracts the useful information from the Big Data.

On the other hand, the load balancer balances the processing power by equal distribution of the real-time data to the servers. The filtration and load-balancing server not only filters and balances the load, but it is also used to enhance then system efficiency.

The proposed architecture and the algorithms are implemented in Hadoop using Map Reduce programming by applying remote sensing earth observatory data.

The proposed architecture is composed of three major units, such as 1) RSDU; 2) DPU; and0 DADU. These units implement algorithms for each level of the architecture depending on the required analysis.

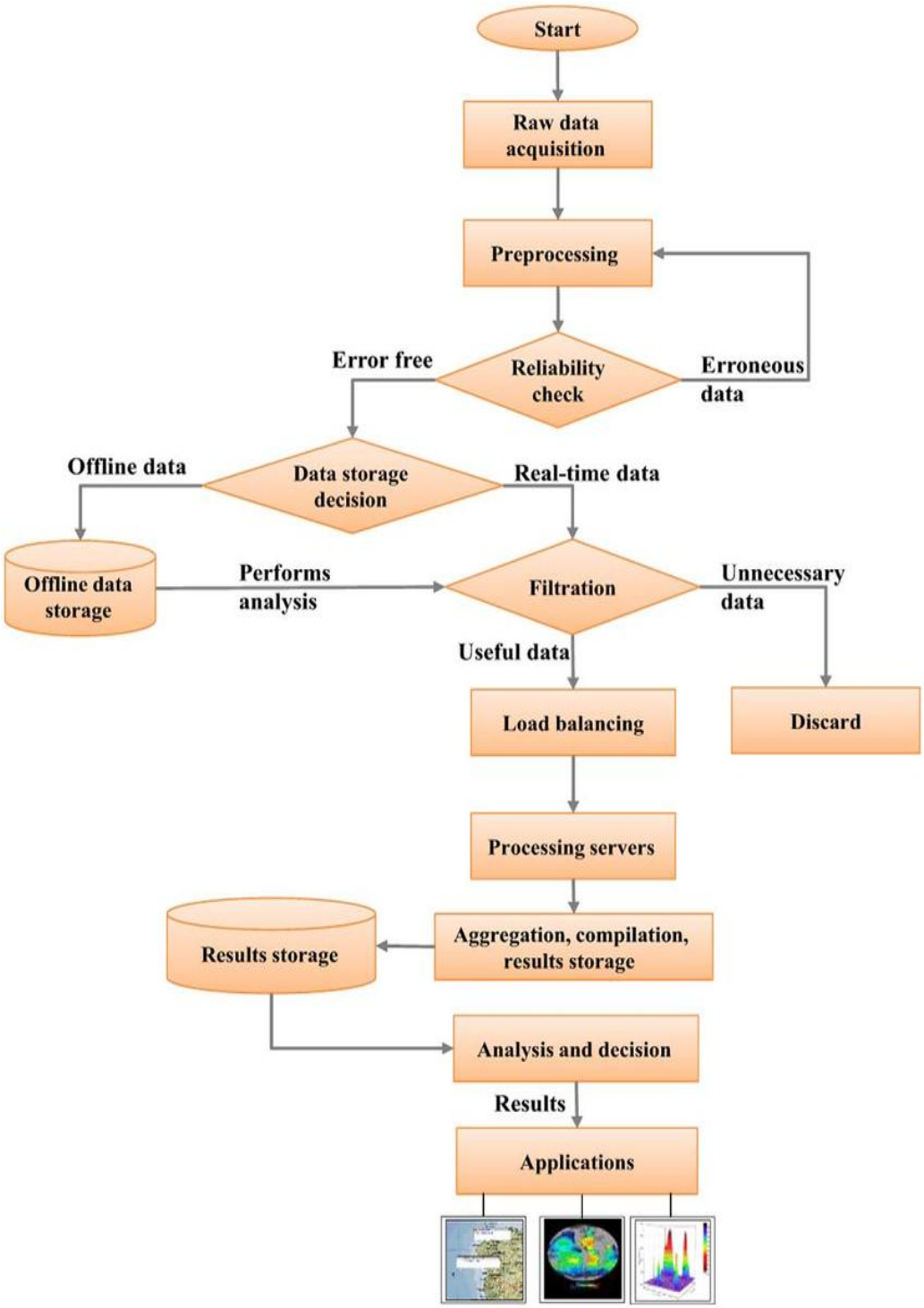
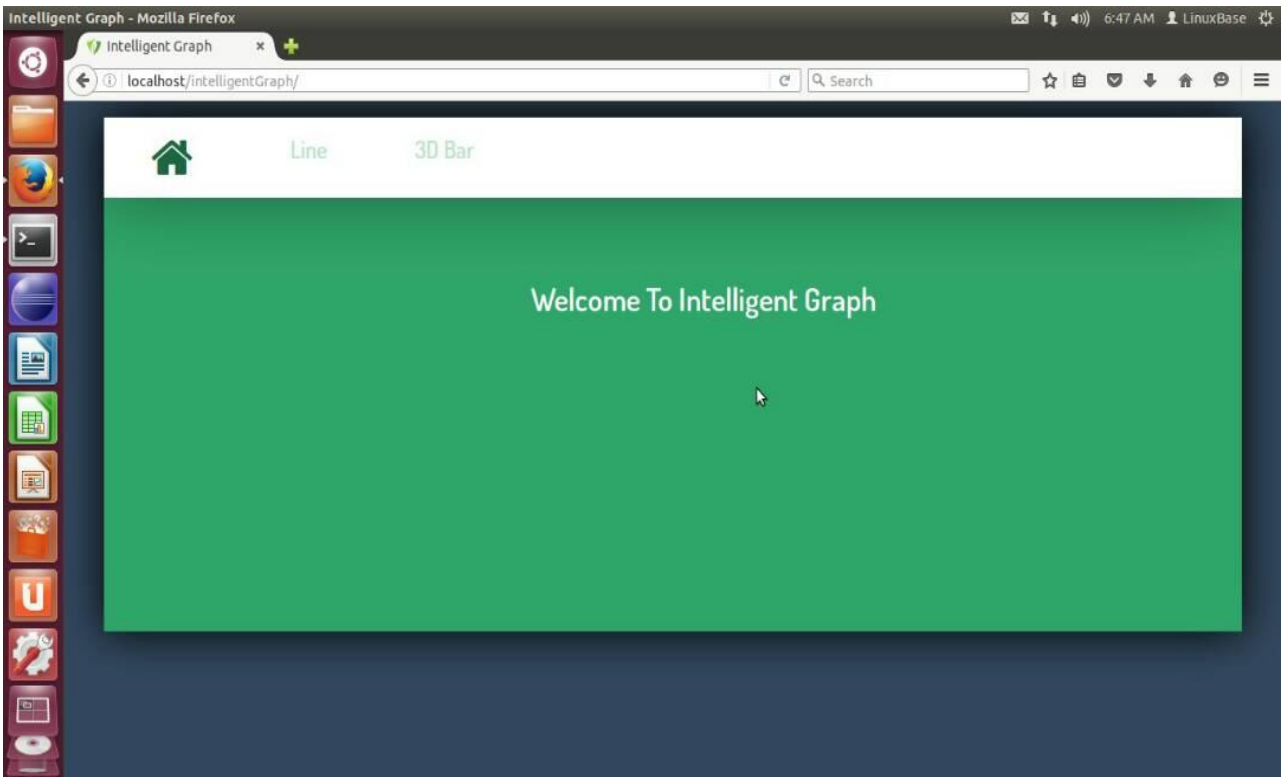
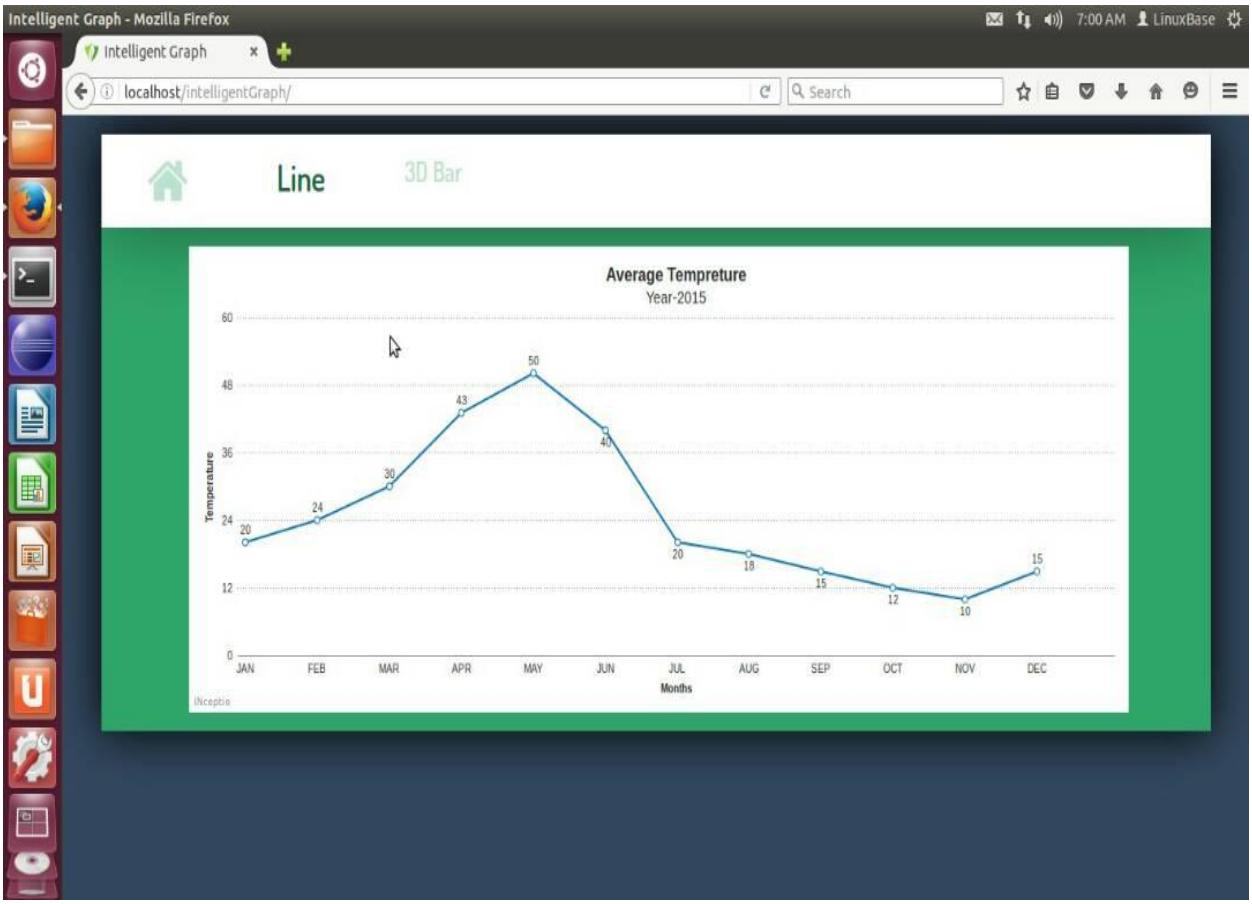
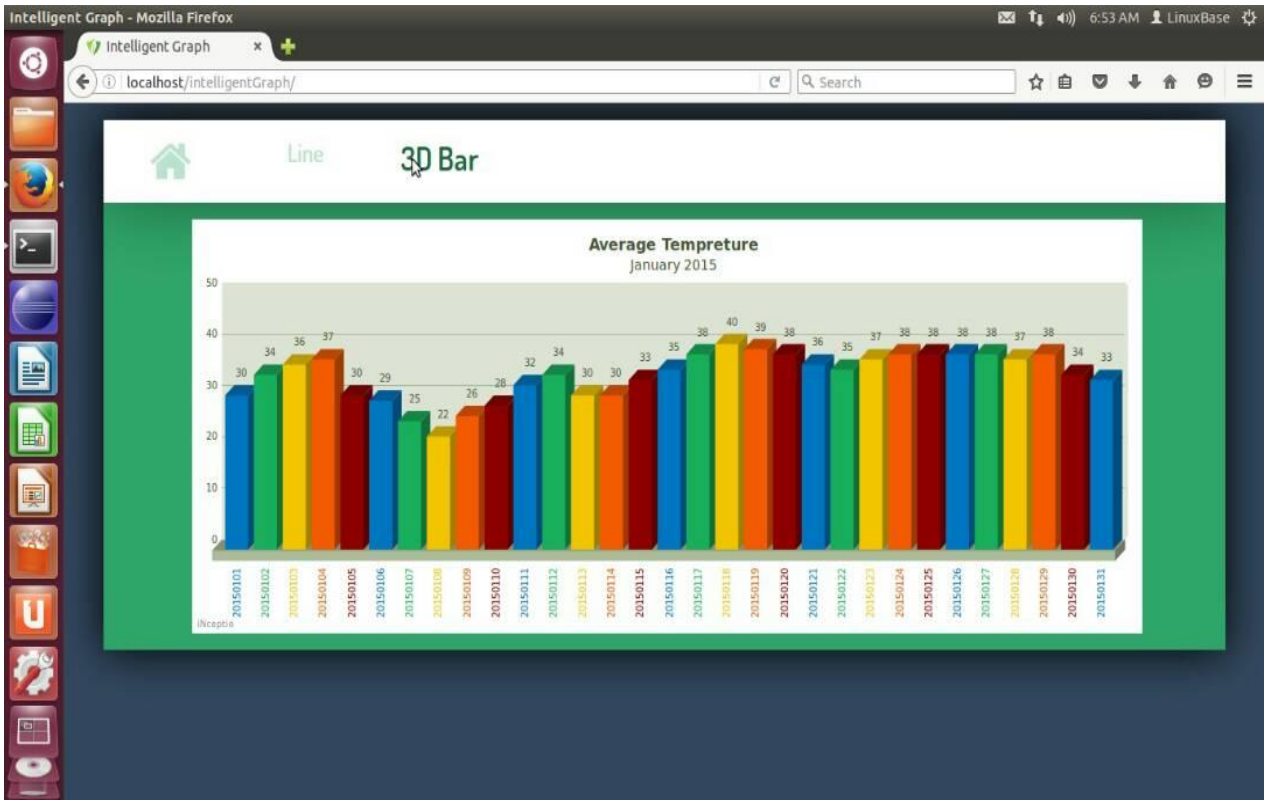


Fig.1 System Architecture

V. Snapshots







VI. Conclusion and Future Work

In this paper, we introduced architecture for real-time Big Data analysis for the remote sensing application. The suggested the architecture efficiently and analyzed the real-time and offline remote sensing Big Data for decision-making. The proposed architecture is designed of three major units, such as 1) RSDU; 2) DPU; and 3) DADU. These units implement algorithms for each stage of architecture depending on the required analysis. The architecture of the real-time Big is generic (application independent) that is used for any type of remote sensing Big Data analysis. Afterwords, the capabilities of the parallel processing, filtering and dividing of only useful information are performed by discarding all other extra data. These processes make a better selection for the real-time remote sensing Big Data analysis. The algorithms proposed in this paper for each unit and subunits are used to analyze remote sensing data sets, which helps in the better understanding of land and the sea area. The architecture proposed welcomes researchers and organizations for any type of remote sensory Big Data analysis by developing algorithms for each level of the architecture depending on their analysis requirement.
For future work, we are planning to extend the proposed architecture to make it compatible for Big Data analysis for all applications, e.g., sensors and social networking. We are also planning to use the proposed architecture to perform complex analysis on earth observatory data for decision making at realtime, such as earthquake prediction, Tsunami prediction, fire detection, etc.

VII. ACKNOWLEDGMENT

We might want to thank the analysts and also distributers for making their assets accessible. We additionally appreciative to commentator for their significant recommendations furthermore thank the school powers for giving the obliged base and backing.

VIII. References

[1] M Mazhar, U Rathore, Anand Paul, A Ahmad, Bo-Wei Chen, Bormin Huang, and Wen Ji, "Real-Time Big Data Analytical Architecture for Remote Sensing Application", IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, 2015.

[2] D. Agrawal, S. Das, and A. E. Abbadi, "Big Data and cloud computing: Current state and future opportunities," in *Proc. Int. Conf. Extending Database Technol. (EDBT)*, 2011, pp. 530–533.

[3] S. Marchal, X. Jiang, R. State, and T. Engel, "A Big Data architecture for large scale security monitoring," in *Proc. IEEE Int. Congr. Big Data*, 2014, pp. 56–63.

[4] X. Yi, F. Liu, J. Liu, and H. Jin, "Building a network highway for Big Data: Architecture and challenges," *IEEE Netw.*, vol. 28, no. 4, pp. 5–13, Jul./Aug. 2014.

[5] E. Christophe, J. Michel, and J. Inglada, "Remote sensing processing: From multicore to GPU," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 4, no. 3, pp. 643–652, Aug. 2011.

[6] Y.Wang *et al.*, "Using a remote sensing driven model to analyze effect of land use on soil moisture in the Weihe River Basin, China," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 9, pp. 3892–3902, Sep. 2014.

[7] K. Michael and K. W. Miller, "Big Data: New opportunities and new challenges [guest editors' introduction]," *IEEE Comput.*, vol. 46, no. 6, pp. 22–24, Jun. 2013.

[8] C. Eaton, D. Deroos, T. Deutsch, G. Lapis, and P. C. Zikopoulos, *Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data*. New York, NY, USA: Mc Graw-Hill, 2012.

[9] R. D. Schneider, *Hadoop for Dummies Special Edition*. Hoboken, NJ, USA: Wiley, 2012.

