

IoT Based Atmosphere Monitoring System using Hadoop Map Reduce Paradigm

K. Purushotam Naidu, P. Krishna Subba Rao, MHM Krishna Prasad

Abstract: In today's world there are a host of serious environmental problems, and air pollution is one of the top causes of many chronic respiratory diseases with accelerated aging of lungs and possibly cancer. Air pollution is the most widespread pollution which is inevitable, since air being an ever pervading medium and carrier, which can transfer the pollutants very fast in no time. It is generally caused due to pollutants like carbon dioxide, carbon monoxide, nitrogen oxides, sulfur dioxide that are released especially when fuels are burnt. This paper aims at designing an Internet of Things (IoT) based atmosphere monitoring system using Hadoop ecosystem. Internet of things (IoT) provides a platform that allows devices to be connected, sensed and be controlled remotely across a network infrastructure. Hadoop is an open source framework which is used to store and process large amount of data using commodity hardware. The key feature of this paper is to measure the level of air contaminants, temperature in the atmosphere using sensors, and also store the sensor data into the cloud server that is Adafruit. Thereby, we redirect the data into Hadoop Distributed File System (HDFS) and process the sensor data to list out the areas with highest pollution level and temperature.

Key words: IoT, sensors, Hadoop, HDFS, Adafruit

I. INTRODUCTION

Environmental pollution is one of the most censorious threat that is being faced by our planet in the current day scenario. This is the issues of global concern which is seen commonly in all the countries, including the wealthiest countries in the world irrespective of their developmental statuses. Environmental pollution is caused when humans introduce contaminants in the environment through their activities which leads to the disruption of the routine processes, causing irrevocable changes in the environment. Pollutants are the substances which occur in the nature or created because of extraneous human activities. The environmental pollution may be classified into six different types based on the pollutants and the pollution caused in the components of the environment. They are air, water, soil/land, noise, radioactive and thermal pollutions. Out of these air pollution is a most serious problem in the present day scenario all over the world especially in the hi-tech cities because of huge level of industrialization. Now-a-days due to the increase in population, the usage of vehicles also increased gradually, resulting in increase of pollution by harmful gases like carbon monoxide that are released due to the combustion of petrol and diesel.

Revised Manuscript Received on February 29, 2020.

Mr K. Purushotam Naidu, Assistant Professor, Department of Computer Science and Engineering, GVP College of Engineering, JNTUK University, Andhra Pradesh.

Prof. Dr. Pulugurta Krishna Subba Rao, Professor, Department of Computer Science and Engineering and Head of the Department, Gayatri Vidya Parishad College of Engineering (Autonomous), Visakhapatnam, Andhra Pradesh,

Dr Munaga HM Krishna Prasad, Professor, Department of Computer Science and Engineering and Vice-Principal, University College of Engineering Kakinada (Autonomous), JNTUK, Andhra Pradesh.

As a result of the air pollution many green house gases are releasing into the atmosphere leading to the global warming. Global warming is defined as an increase in the average temperature of the Earth due to air pollutants, which collect sunlight and radiation and produce green house effect. As a result of these many health issues have been arising causing severe damage to the animals including human beings. Many respiratory diseases like asthma, chronic obstructive pulmonary disease, pulmonary fibrosis, and pneumonia and lung cancer are the ramifications of inhaling these harmful gases.

So in order to reduce this pollutant level in the atmosphere we have to keep the complete track of the pollutant level. In recent years Internet of Things (IoT) has begun to play a crucial role in our daily lives with the ability to modify the environment around us. IoT platform [5] is a multi-layer technology which offers a platform for sensors and devices to communicate consistently within a smart environment and enables the sharing of information across various platforms in a convenient manner. The adaption of various wireless technologies in the recent days places IoT as the next revolutionary technology by benefiting from all the opportunities offered by the Internet technology. Thus this paper aims to apply IoT for both monitoring and sensing the pollutant level in the atmosphere. In order to display the data in the real time online we use Adafruit IO for visualizing multiple feeds of data.

In such visualization, a Wi-Fi enabled system on chip module ESP8266 is used [1]. It is capable of either hosting an application or offloading all Wi-Fi networking functions from another application processor. A basic ultra low-cost digital temperature and humidity sensor DHT 11 uses a thermistor and a capacitive humidity sensor and does some analog to digital conversion and finally gives a digital signal on the data pin. It calculates relative humidity by measuring the electrical resistance between two electrodes. The MQ-2 gas sensor module is useful for detecting the gas leakage. It has the capability of detecting LPG, hydrogen, smoke, i-butane, methane and propane. It measures and indicates the concentration of certain gases in air hence these are employed in numerous industries and also in monitoring the atmosphere for detecting the release of these harmful gases.

In this paper IoT has been applied for connecting various physical devices like DHT 11 and MQ-2 gas sensors which are connected to ESP8266 module and uses Adafruit cloud server for data storage [2-4].

IoT Based Atmosphere Monitoring System using Hadoop Map Reduce Paradigm

Hadoop is an open source frame work ,which is used to store and process large volume of data on commodity hardware.Mainly hadoop contains two components one is HDFS and another one is map reduce.HDFS is an Hadoop Distributed File System used to store data that is collected from different sources. Map reduce is frame work used to process the data that is stored in HDFS. Hadoop having five basic building blocks.In this first one is secondary name node, it periodically maintains the log information of name node. Second one is name node,which maintains meta data for HDFS.Here meta data holds block locations of data nodes,replication factor,working conditions of data nodes etc.Third one is data node, it stores the actual data that is used to process.fourth one is job tracker,it tries to assign tasks to task tracker on the data node.Fifth one is Task tracker that accepts tasks from job tracker and periodically send heart beta messages to job tracker to notify that it is alive.

II. LITERATURE SURVEY

In the earlier systems for determining the amount of contaminants present in the air we use many number of sensors for determining the amount of each contaminant. Using large number of sensors gradually increases the cost of implementation by increasing the cost of equipment. Due to the usage of large number of sensors the circuit becomes very difficult to handle. Moreover usage of too many sensors increases the amount of power consumed. In the earlier systems very less number of calculations are performed and the time consumed in performing these calculations is also very high. If the amount of data to be computed is very large, the earlier systems can cause great number of errors. For the air pollution monitoring, earlier systems use Fourier Transform Infrared instruments (FTIR) which are not user friendly. Continuous monitoring of pollution levels is not observed in the earlier system that is pollution levels are checked only at a particular interval of time. Therefore small fluctuations in the level of pollution cannot be computed accurately. The gas chromatographs, mass spectrometers and other instruments used are not very accurate. In the earlier systems the space complexity is very high because the cloud storage concept using Adafruit was not introduced.

Author [2] says that devices are connected through IoT and they produce large volumes of data. This paper presents Big Data as a solution to handle these large amounts of data and applications of IoT using Big Data.Author [3] says that a city gets evolved into a smart city with the help of ICT and IoT. This paper aims to collect information from smart city sensors and transform this information into data forms like RDF and JSON. XML data model helps in this data transformation. Author [4] provided a method to overcome the challenges faced by the farmers in reducing the water consumption and providing better irrigation schedules. This paper focuses on wireless sensor network for collecting the data and discusses about its structure and presents the graphical representation of the received data.Author [6] proposed an online monitoring system for monitoring the crop growth, soil moisture, agricultural natural disasters.These systems are based on cloud

computing.Through these monitoring systems, the Internet of things intelligent gateway, cloud based research and construction of large data analysis and data mining techniques are used to analyze large amount data[10-13].Author [7] says that monitoring the temperature and moisture of single soil layer is not sufficient to represent growth around the crop. This paper aims at monitoring system for multilayer soil temperature and moisture based on sensor networks.Author [8] proposed a method for collecting and managing sensors data in a smart building which operates in IoT environment. Cloud server is used to store the data produced from sensors in smart building. This data can be managed by a remote device. Installing multiple sensors could lead us in efficient smart building and thus in a Green Smart Building.The proposed systems are implemented in a simulation environment of Cooja Contiki.

Author [9] proposed a system to overcome the problems an ordinary citizen faces. This paper presents the Smart Digital City architecture and implementation model to design the system. This system uses IoT devices to collect the real time data and Big Data to process the real time data in an effective manner.This system also helps the government to take smart decisions using real time data.

III. PROPOSED SYSTEM

In Proposed System we are using arduino node MCU ESP8266 module ,temperature sensor and gas sensor.Both gas and temperature sensors connected to node mcu,which collects data from sensors and stores into cloud.To store data into cloud we use Adafruit IO.It is an cloud service that means it allows you to chart,graph,gauge,log and displays data from anywhere in the world from the internet. In this we are collected data from different locations and stored it into adafruit. Once data stored into cloud service ,the big data analytics used to store and process data by using hadoop map reduce paradigm.Here we are collecting temperatre,humidity,gas and location etc. from the cloud service then store it into HDFS for processing.HDFS is an hadoop Distributed File System for storing structured,unstructured and semi structured data.The data that is stored in HDFS consists of set of records from each record we extract location,Temperature and gas value using hadoop mapper and reduce paradigm.

In this the temperature, humidity and pollution levels are streamed continuously and computations are performed on the data obtained. In this system instead of using a large number of sensors for computing a single contaminant in the air we use a single sensor which does all the work of multiple sensors. This causes reduction in the cost of implementation by reducing the cost of equipment. Usage of a single sensor also reduces the power consumption. Due to the usage of a single sensor the circuit becomes easy to handle. The proposed system is least prone to errors because the huge amount of data obtained is analyzed using big data analytics. Unlike the earlier systems pollution levels are monitored continuously so that small fluctuations in the level of pollution can be computed accurately. In the proposed system noxious gases like benzene can be detected using the sensor.

In the following fig.1 shows that ESP 8266 Node MCU Wifi module reads data from DH11 Sensor and gas sensor that is stored into Ada fruit Cloud service. From this Sever transform data into HDFS for processing using Map Reduce. After Applying Algorithm1 and Algorithm2 the results are shown those are locations with maximum pollution and maximum tempetrature.

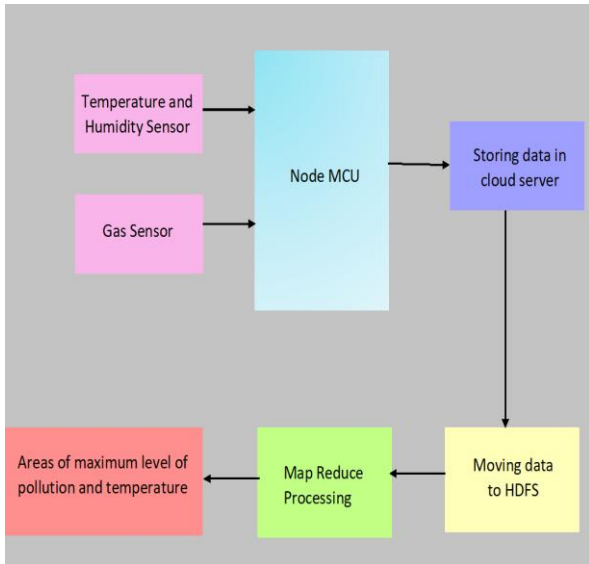


Fig 1:General Structure of pollution Monitoring System

The following algorithms describe how to get maximum temperature and gas at each location using hadoop framework. In map reduce framework both mapper and reducer takes <key,value> pair as input and output. Mapper and reducer having input and output types.

Map input:< key:record id,value:record>

Map output:< key:location,value:record>

Algorithm MaxMapper< LongWritable key, Text value, Context context>

```
//read each record one by one
1 record:=value.toString();//if record length
  10 then extract location from index 7 to 9
2 if(record.length()==10){
3 key1.append(record.subsrting(7,10));
  //load map output
4 context.write(newText(key1.toString()),new
  Text(record.toString()));//if record length 11 then
  extract location from index 8 to 10
5 }else if(record.length()==11){
6 key1.append(record.substring(8,11));
  //load map output to reducer
7context.write(newText(key1.toString()),new
  Text(record.toString()));
End Algorithm
```

Algorithm 1:Max Mapper

The input to Mapper Algorithm is record from input file. Each record length may be 10 or 11 that depend on the gas. If the gas value is three digit number then record length is 10 otherwise if gas value is four digit number then record length is 11. So length of record is depends on gas value.

Reducer input:<key:location,value:record>

Output:<key:location,value:(maximum temperature,maximum gas)>

Algorithm MaxReducer< Text key, Iterable<Text> values, Context context>

```
1 max=Integer.MIN_VALUE;
2 max2=Integer.MIN_VALUE;
3 String loc="";
4 String loc1="";
//for each value that may be temperature or gas
5 for (Text val : values){
6 String val1=val.toString();
//if length 10 then identify max temperature and gas value
7 if(val1.length()==10){
//extract temperature
8 val2=Integer.parseInt(val1.substring(0,2));
9 max=Math.max(max, val2);
10 if (max==val2){
//extract location
11 loc=val1.substring(7, 10);}
//extract gas value
12 val3=Integer.parseInt(val1.substring(4,7));
13 max2=Math.max(max2, val3);
14 if (max2==val3){
15 loc1=val1.substring(7, 10);}}
16 else if(val1.length()==11){
//extract temperature
17 val2=Integer.parseInt(val1.substring(0,2));
18 max=Math.max(max, val2);
19 if (max==val2){
//extract location
loc=val1.substring(8, 11);}
20 val3=Integer.parseInt(val1.substring(4,8));
21 max2=Math.max(max2, val3);
22 if (max2==val3){
//extract gas value
23 loc1=val1.substring(8, 11);}}
24 String str="";
//concate temperature and gas value
25 str=max+" "+max2;
//reducer output key is location and value is temperature
and gas value
26 context.write(new Text(loc.toString()), new
  Text(str.toString())); }
27 end Algorithm
```

Algorithm 2:Max reducer

The output of algorithm1 taken as input to reducer algorithm and output of reducer algorithm is location, maximum temperature and maximum gas. In reducer algorithm from line 5 we iterate each record in which extract maximum temperature and gas at particular location from line 6 to line 25 then finally store the result in HDFS from line 26.

IV. SYSTEM DESIGN AND IMPLEMENTATION

The DHT 11 sensor is used to read humidity and temperature data. It's perfect for remote weather stations, environmental control systems and farms or garden monitoring systems.

IoT Based Atmosphere Monitoring System using Hadoop Map Reduce Paradigm

The Vcc of the DHT 11 sensor is connected to +5V of the Arduino. The ground (GND) of the DHT 11 is connected to the GND of Arduino. The data pin or signal pin is connected to the pin 2 of the Arduino as shown in the Fig1.

The Grove-Gas sensor (MQ-2) module is useful for gas leakage detection. Gases like H₂, LPG, CH₄, CO, alcohol, propane are detected using MQ-2 sensor. Measurement can be taken in an efficient way due to its high sensitivity and fast response time. Potentiometer can be used to adjust the sensitivity of the sensor. The Vcc of the MQ-2 is connected to +5V of Arduino. The GND is connected to the GND of the Arduino. The signal pin is connected to the A0 pin of the Arduino as shown in the Fig.2. The core of the Adafruit IO system is feeds. The feed also contains the sensor data values that are pushed from our device to the Adafruit IO.

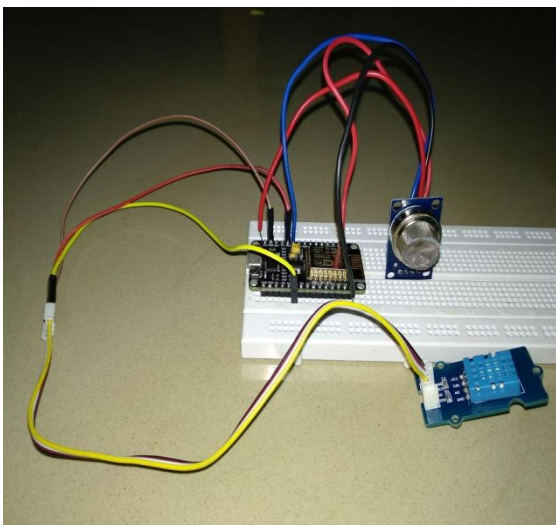


Fig.2. System architecture and implementation model

In Atmosphere monitoring system, we deal with the sensed data sensed by the two sensors that is stored in the cloud server. The following functions are being performed by the proposed system in sequence.

1. The required atmospheric parameters are measured.
2. Transmit the data to the surface location.
3. Process and store the measured data.
4. Reports are created by analyzing the data.

This work mainly aims to design and implement systems with sensors in the atmosphere and data management and storage using the cloud server Adafruit. The extraction of important and useful knowledge from a large data on temperature and humidity, are obtained with IoT's . An open-source electronics platform based on easy-to-use hardware and software is Arduino. The Arduino board is connected to the computer using an USB, where it connects with the Arduino Integrated Development Environment (IDE). The Arduino code is written by the user in the IDE and then uploads it to the microcontroller which executes the code, there by interacting with the sensors that are connected to the ESP8266 wifi module.

The algorithm for the arduino code is as follows.

```
#include libraries //includes all the necessary libraries
#define macros //define the pin numbers to which the
sensors are connected
void setup ()
{
```

```
// initialize all variables, pin modes, start using
libraries etc
//this function will run only once after each
powerup or reset of arduino board
}
void loop()
{
//this is repeated continuously because it is
called by the hidden loop that controls the execution of the
arduino program.
}
```

In general the structure of arduino code as shown above. This structure is similar to C++ Programming. The first part contains header files, the setup function is a predefined function that is used to initialize all variables, pin modes, start using libraries etc. This function will run only once after each power up or reset of arduino board. The actual logic is written in predefined loop function. This is repeated continuously because it is called by the hidden loop that controls the execution of the arduino program.



Fig.3: Ada fruit cloud data

The above Fig.3 shows that the data is stored into Ada fruit cloud service. In this value field contains temperature, humidity, pollution value and location. The length of this filed is 10 or 11. if pollution value is 3 digit number then length is 10 otherwise if pollution value is 4 digit number then length is 11 this is already discussed in above Algorithm 1 and Algorithm 2.

```
{
  "id": "2018-03-20T15:57:39.834Z",
  ("id": "00T7M6973782VMS6F8JGQ8R", "value": "3221180um", "feed_id": 745203, "create
  d_at": "2018-03-20T15:57:42.403Z",
  ("id": "00T7M696WZ7C24C26WPKXXR", "value": "3221180um", "feed_id": 745203, "create
  d_at": "2018-03-20T15:57:44.955Z",
  ("id": "00T7M626A9E9340WQ0VEJ", "value": "3221180um", "feed_id": 745203, "create
  d_at": "2018-03-20T15:57:47.495Z",
  ("id": "00T7M6AKAVDQ40SD0HRYECK", "value": "3221180um", "feed_id": 745203, "create
  d_at": "2018-03-20T15:57:50.643Z",
  ("id": "00T7M6Z3V5Z6P91HE1CSP259G", "value": "3221180um", "feed_id": 745203, "create
  d_at": "2018-03-20T15:57:52.696Z",
  ("id": "00T7M626A9E9340WQ0VEJ", "value": "3221180um", "feed_id": 745203, "create
  d_at": "2018-03-20T15:57:55.143Z",
  ("id": "00T7M6NG01W3M21Y0YAD096", "value": "3221180um", "feed_id": 745203, "create
  d_at": "2018-03-20T15:57:57.683Z",
  ("id": "00T7M61E61N1NE41A5EP2K220G", "value": "3221180um", "feed_id": 745203, "create
  d_at": "2018-03-20T15:58:00.211Z")
}
```

Fig 4: Input data set collected from Ada fruit cloud service

Once the data is stored into cloud ,that can be transformed into HDFS to process using Hadoop map reduce paradigm. Sample structure of the data that is stored in HDFS is shown in Fig.4.Each record is enclosed with in {}.In this the field contains ID ,second filed contains the actual required data to process and third field contains feed-id and last filed contains time.



Fig.5:Displaying basic information to the user

Fig.5 shows the basic information to the user that having current value and get results tabs.if we click on the get results tab it displays current current temperature and pollution level to the user that is shown in below Fig.6.

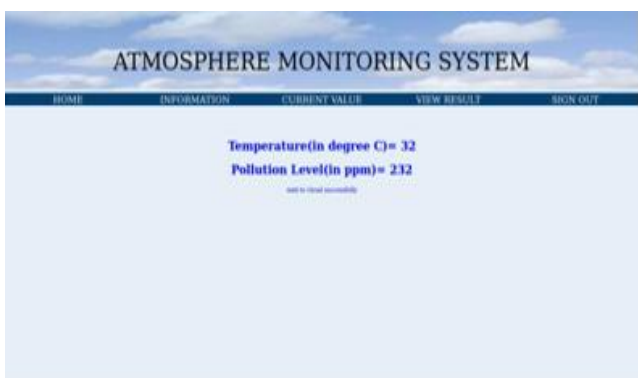


Fig.6: Live Streaming Value

If we click on view results button is shows location with maximum temperature and maximum pollution. These results are displayed by running map reduce task in the back ground and store the results into HDFS. From the HDFS the results are displayed into web browser as shown in Fig.7.



Fig.7: User views the max temperature and max pollution

The results from Table 1 categorized as Highly polluted areas,ready to pollute areas and areas with clean air. We are collected pollution, temperature values from different areas then stored into adafuit IO cloud from this we are processed using Hadoop map reduce paradigm. From the given results Gajuwaka having highly polluted area, Nad area is ready to pollute and Kommadi and Simhachalam are clean areas. As considered above 1000 is highly polluted,800-1000 is ready to pollute and less than 800 is clean air as threshold values.

Table 1:Pollution values at different locations

LOCATION	POLLUTION VALUE	OUTCOME
GAJUWAKA	1003	HIGHLY POLLUTED
NAD	854	READY TO POLLUTE
SIMHALAM	651	CLEAN AIR
KOMMADI	234	CLEAN AIR

V. CONCLUSION

In recent years Internet of Things (IoT) has begun to play a crucial role in our daily lives with the ability to modify the environment around us.Thus this paper aims to apply IoT for both monitoring and sensing the pollutant level in the atmosphere. In order to display the data in the real time online we use Adafuit IO for visualizing multiple feeds of data. After data stored in Adafuit IO,Hadoop framework helps to process that data to monitor the pollution levels in the atmosphere.

This paper Combined Internet of Things, Adafuit cloud server and sensor technologies with an aim to make use of the data collected by them to monitor the pollution levels in the atmosphere. Therefore, we have to put in efforts to update our lives to the Internet of Things technology with respect to both hardware and software.

REFERENCES

1. Alansari, Z., Soomro, S., Belgaum, M. R., & Shamshirband, S. (2018). The Rise of Internet of Things (IoT) in Big Healthcare Data: Review and Open Research Issues. In *Progress in Advanced Computing and Intelligent Engineering* (pp. 675-685). Springer, Singapore.
2. R. Y. Zhong, X. Xu, and L. H. Wang, "IoT-enabled Smart Factory Visibility and Traceability using Laser-scanners," *Procedia manufacturing*, 2017.
3. R. Y. Zhong, L. H. Wang, and X. Xu, "An IoT-enabled Real-time Machine Status Monitoring Approach for Cloud Manufacturing," *Procedia CIRP*, vol. 63, pp. 709-714, 2017.
4. C. Stergiou, K. E. Psannis, "Efficient and Secure Big Data delivery in Cloud Computing", Springer, Multimedia Tools and Applications, pp.1-20, April 2017.
5. Tiainen, P., New opportunities in electrical engineering as a result of the emergence of the Internet of Things. 2016.
6. Al Nuaimi, E., et al., Applications of big data to smart cities. *Journal of Internet Services and Applications*, 2015. 6
7. Ahvenniemi, H., Huovila, A., Pinto-Seppä, I., & Airaksinen, M. (2017). What are the differences between sustainable and smart cities? *Cities*, 60, 234-245.
8. R. Y. Zhong, C. Xu, C. Chen, and G. Q. Huang, "Big Data Analytics for Physical Internet-based intelligent manufacturing shop floors," *International Journal of Production Research*, vol. 55, pp. 2610-2621, 2017.
9. A.P. Plageras, K.E. Psannis, C. Stergiou, H. Wang, B.B. Gupta, Efficient IoT-based sensor BIG Data collection-processing and analysis in Smart Buildings, *Future Generation Computer Systems* (2017). <https://doi.org/10.1016/j.future.2017.09.082>
10. C. Stergiou, K. E. Psannis, "Recent advances delivered by Mobile Cloud Computing and Internet of Things for Big Data applications: a survey", *International Journal of Network Management*, pp. 1-12, May 2016.
11. Yunchuan Sun, Houbing Song, Antonio j.jara, Rongfang Bie "IOT and Big Data Analytics for Smart and Connected Communities" *IEEE journals and magazines*, 2016, volume: 4.
12. M. M. U. Rathore, A. Paul, A. Ahmad, B. W. Chen, B. Huang and W. Ji, "Real-Time Big Data Analytical Architecture for Remote Sensing Application," in *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 10, pp. 4610-4621, Oct. 2015.
13. Marz, N., & Warren, J. (2012). *Big data: Principles and best practices of scalable real time data systems*. Manning: MEAP edition.

AUTHORS PROFILE



Mr K.Purushotam Naidu, is currently Working as Assistant Professor, Department of Computer Science and Engineering, GVP College of Engineering, JNTUK University, Andhra Pradesh. He did his B.E. from SVHCE, Nagarjuna University, Guntur, M.Tech from Andhara University Visakhapatnam and pursuing Ph.D

from JNTUK University. His research interests include data mining, BigData Analytics, Internet of Things and Machine Learning.



Prof. Dr. Pulugurta Krishna Subba Rao, is currently Full Professor, Department of Computer Science and Engineering and Head of the Department, Gayatri Vidya Parishad College of Engineering (Autonomous), Visakhapatnam, Andhra Pradesh, India. He did his Ph.D in Computer Science & Engineering from Acharya Nagarjuna University Guntur India. He

has about 30+ research papers in various International Journals and Conferences, and attended many national and international conferences in India and abroad. He is a member of IEEE and he is an active member of the board of reviewers in various International Journals and Conferences. His research interests include data mining, Bio Informatics, Network Security and Internet of Things. He received prestigious the Institution of Engineers (India) Membership certificate of Fellow (F-12593-9) on the fourteenth day of August two zero one nine



Dr Munaga HM Krishna Prasad is currently Full Professor, Department of Computer Science and Engineering and Vice-Principal, University College of Engineering Kakinada (Autonomous), JNTUK, Andhra Pradesh. He did his B.E. from C.B.I.T, Osmania University, Hyderabad, M.Tech. and Ph.D. Computer Science and Engineering (in general) and (in specific data mining, Thesis Title: "Trajectory Clustering Algorithm: A Data Mining Solution to Industrial Problems – Critical Analysis and Evaluation") from JNTU, Hyderabad. Dr Munaga successfully completed a two year MIUR fellowship (Jan 2007 – Dec 08) at University of Udine, Udine, Italy. He has about 50+ research papers in various International Journals and Conferences, and attended many national and international conferences in India and abroad. He is a member of Association for Computing Machinery (ACM), Computer Society Of India (CSI), ISTE and IAENG (Germany) is an active member of the board of reviewers in various International Journals and Conferences. His research interests include data mining, BigData Analytics and High Performance Computing. He received prestigious STATE TEACHER AWARD from the Hon'ble Chief Minister, Govt. of Andhra Pradesh on 5th Sept. 2019.