Low cost remote heart beat monitoring and abnormality prediction using Hadoop cluster.

Tej Sai Ram J, Vivek Vardhan R, Venkata Sridhar T

Abstract— In modern life style humans have no time to monitor and properly maintain the health records.According to the survey of WHO an estimated 17.7 million people died of CVDs in recent years, representing 31%of all global deaths. Of these deaths, an estimated 7.4 million were due to coronary heart disease and 6.7 million were due to stroke. Hence to help the remotely located people we tried to aim in this paper is to develop a low cost portable heart beat monitoring and abnormality prediction system using apache hadoop. The device is designed with three sensing electrodes, ECG sensor and NodeMCU. This makes it compact and compatible for our daily usage. The utility of such devices will increase the data sets exponentially, such huge amounts of data is best processed using parallel computing. In our system we are using Hadoop multimode cluster which incorporates parallel computing and data replication for high processing speed and prevention of data loss.

Index Terms— Apache Hadoop, Data Replication, Multinode cluster, NodeMCU, Parallel computing, Telecardiology.

1 INTRODUCTION

Humans are most commonly affected by cardiac diseases nowadays. Many people are suffering from cardiac diseases due to stress and some of them had to take up coronary artery bypass surgical treatment for their cardiac problem. . To prevent this unwanted heart attack, early detection and timely treatment of arrhythmia is necessary. This would slow down the progression of heart failure and prevent the loss of life [1].

The best non invasive method to observe cardiac condition is ECG signal. ECG signal can be acquired using AD8232. This is used for biopotential measurement in our body. It is designed to extract, amplify, and filter small biopotential signals in the presence of noisy conditions, such as those created by motion or remote electrode placement. This design allows an embedded microcontroller to easily acquire the output signal.

The graphical recording of the electrical signals generated by the heart is known as ECG which provides diagnostic information about the cardiac condition of a patient. P wave is the electrical positive signal in the normal ECG. . It has positive polarity and duration is less than 120 milliseconds. The largest part of the ECG signal is the QRS complex which is the result of contraction of the ventricles where:

a. Q wave is the first negative or downward deflection.
b. R wave is always the first positive deflection,
c. S wave is the negative deflection followed by the R wave [2].

1.1 Characteristics of Big Data

There are 4V's in bigdata. They are

Volume

Volume indicates the amounts of data generated every second. It mainly referred as quantity of data. The data available through social website and sensor networks going to cross from petabytes to zetabytes.

Variety

Variety refers to the different types of data. In the past, researchers only focused on structured data and data are arranged in tabular form like financial data, Structured, semi-structured, unstructured; Text, image, audio, video, record.

Velocity

Velocity Refers to the speed of data processed. This is a concept which indicates the speed at which the data generated and become historical. Big data is able to handle all types of data.

Variability

It describes the amount of variance used in data that kept within the data bank and refers how they are spread out or closely clustered within the data set [4].

1.2 Telecardiology

Telecardiology can be defined as the monitoring or diagnosis of cardiac activities at a distance via telecommunication technology. ECG and imaging-based echocardiography (ECHO) are tools most often applied in cardiology. ECHO has become a widely-used tool in telecardiology due to its ability to physically evaluate cardiac and vascular anatomical structures and physiological functions, which can affect intervention strategies. The strongest advantage of telecardiology is that it allows timely remote diagnosis by cardiologists and for the provider to evaluate effective therapeutic strategies, especially for rurally located patients where professional cardiologists are not as accessible. Telecardiology lowers the mortality rate for patients with heart attack and can reduce the cost of

- *Tej Sai Ram J is currently pursuing Bachelor of Technology in Electronics and Telecommunication Engineering at IIIT Bhubaneswar, India,PH-09989191414. E-mail: b214028@iiit-bh.ac.in*
- *Vivek Vardhan R is currently pursuing Bachelor of Technology in Electronics and Telecommunication Engineering at IIIT Bhubaneswar, India,PH-07036882943. E-mail: b214044@iiit-bh.ac.in*
- *Venkata Sridhar T is currently working as Assistant Professor in Electronics and Telecommunication Engineering at IIIT Bhubaneswar, India,E-mail: venkatasridhar@iiit-bh.ac.in*

transportation from the home to the emergency setting or unnecessary transfers between hospitals [6].

A critical tool for telecardiology applications is wireless telecommunication such as wifi, which delivers pervasive services with less interruption errors when compared to traditional telephone lines. Therefore, thanks to this technology, individuals residing in rural areas or disparate health care areas around the world will benefit from remote. However, it is possible to achieve ECG monitoring in a home environment without Internet connections using only traditional telephone lines by recording ECG signals as audio input which is the transmitted to a hospital via a fixed phone line or mobile phone [5].

2 System Components

2.1 Node MCU

Node mcu is an open source iot platform which has both a 32 bit processor (Lx106) and a wifi module (esp8266) embedded in it. It has 128kilo bytes memory and 4mega bytes storage. This module is a recent addition to market which has a lot of features in it for optimal cost. We can process sensor data and send it using wifi module to any server. In this paper we are receiving ECG signal from the sensor and sending the required data from node mcu to Java apache mail enterprise server.

2.2 Apache JAMES

Java apache mail enterprise server also known as james is an open source SMTP, POP3 mail transfer agent and NNTP news server written in java. The James project manages the Apache Mailet API which defines matchers and mailets. These allow users to write their own mail-handling code, such as to update a database, build a message archive, or filter spam.

2.3 Apache Flume

Apache Flume is a distributed, reliable, and available service for efficiently collecting, aggregating, and moving large amounts of log data. It has a simple and flexible architecture based on streaming data flows. It is robust and fault tolerant with tunable reliability mechanisms and many failover and recovery mechanisms. It uses a simple extensible data model that allows for online analytic application.

2.4 Apache Kafka

Apache Kafka is an open-source stream processing software platform developed by the Apache Software Foundation written in Scala and Java. Kafka aims to provide a unified, high through put, low latency platform for handling real-time data feeds. Additionally kafka connects to externl systems for data import and export via kafka connect.

2.5 Spark & Scala

Apache Spark is an open-source cluster-computing framework. Spark provides an interface for programming entire clusters with implicit data parallelism and fault tolerance. Apache Spark requires a cluster manager and a distributed storage system. Spark API can use scala, python and java languages.

3 System Architecture

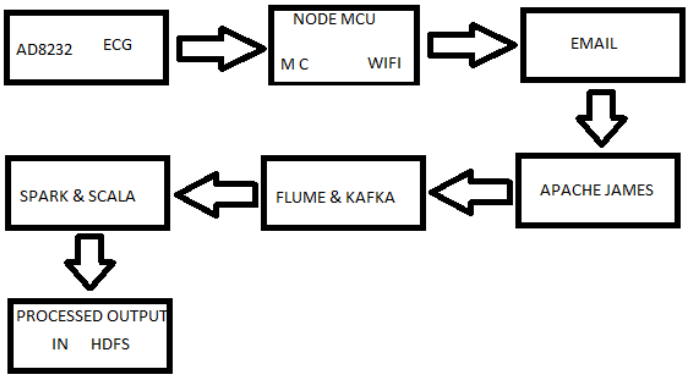


Fig. 1 block diagram of the system

The above figure depicts the process of acquiring the sensor data using ECG sensor and the data is sent to node mcu where the data is preprocessed and that data is sent as an email using the built in wifi module available in node mcu, now the email is sent to apache james, flume is used to collect the data recieved by james and store it in a kafka sink. Now from kafka we use spark to take the data and process it for any abnormalities in the cardiac parameters.

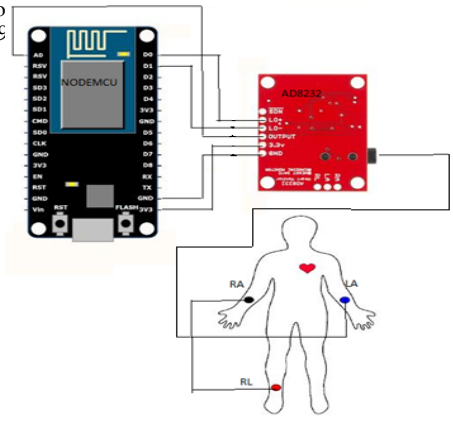
4 Proposed System Features

The device used for ecg data acquisition and transfer is portable and compact. The patients who are unable to be present in the care hospitals either due to economical factor or due to time factor can use this device to continuosly monitor their heart condition and check any abnormalities in its function. The usage of parallel processing in hadoop makes the work done in relatively less time than serial processing. [3]

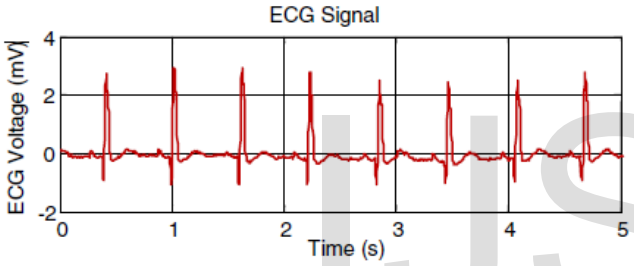
5 Design and Implimentation

5.1 ECG Signal Fetching

In clinical diagnoses involving the ECG signal, it is of the utmost importance that the profile of the signal be as faithfully preserved as possible. The factors affecting the quality of the recorded ECG signal are the skin-electrode-amplifier interface, electrode motion artifact, electrical interference, amplifier CMRR, amplifier frequency response, semiconductor noise generated in the amplifier, and input signal level variation. In this paper, the CMRR is considered.



Since the analog-to-digital converter (ADC) in Atmega 32 microcontroller (MC) can only read the positive part of the signal, a bipolar-to-unipolar circuit is used to shift and amplify the signal above zero. Basically, the bipolar-to-unipolar converter is a non-inverting summing amplifier circuit. Finally, a clamper circuit with 0.7V was added to make sure that the signal is shifted before inserting it into the ADC pin of the MC [7].

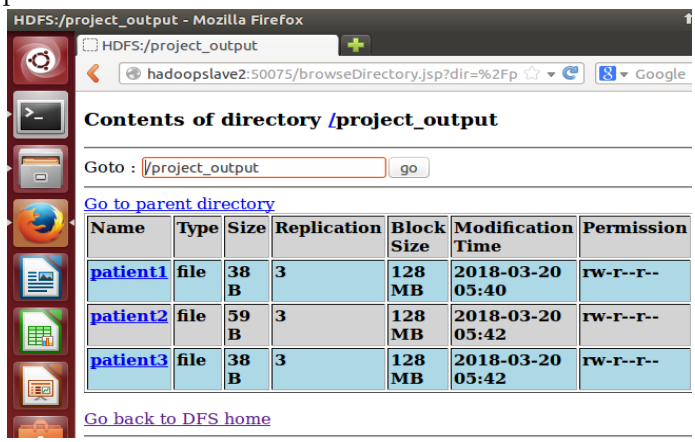


5.2 Hadoop implementation

Apache james, flume, kafka, spark have been setup on the node. We setup Microsoft journal system for sending the mail to james. The email is received at apache james in hadoop, now the flume is configured such that the mail received at james is moved to kafka. Now we process the data received in kafka using spark with scala and send feed back to the patient.

6 Results

The test cases performed in a multimode hadoop environment and their outputs which are accessible from any node placed in different locations.



7 FUTURE SCOPE

The usage of hadoop framework to store and process the data has numerous advantages such as scalability, compatibility with other relational data bases such as sql. So, we can personalize the system according to our needs and we can directly ingest data from sql. The portable device can be manufactured efficiently at low cost and made compatible with other data bases. We can integrate all the health sensor data or any medical data in hadoop and maintain a segregated platform for all the data. The security issues can be resolved by adding authentication for any data transactions through the system.

8 CONCLUSION

Cardiovascular disease is a major health problem in India.In this paper, we proposed a low cost heart beat monitoring and abnormality prediction system using hadoop cluster.The proposed system involves ECG data collection from AD8232 and nodemcu and this data is processed in the hadoop cluster using spark with scala.With a view of developing a continuous monitoring of remotely located people healtdata, this system can handle large amounts of data efficiently.

REFERENCES

[1] S. Jayalalitha, D. Susan, Shalini Kumari and B.Archana, "K-nearest Neighbour method of Analysing the ECG Signal(To find out the Different Disorders Related to Heart)", Journal of Applied Sciences, 14: 1628-1632

[2] Malgina, O., Milenkovic, J. ; Plesnik, E. ; Zajc, M. ; Tasic, J.F., "ECG Signal Feature Extraction and Classification Based on R Peaks Detection in the Phase Space", pp. (381-384), February 2011.

[3] Xuecan Huang, Huhe Dai, and Ye Li*, "The Parallel Processing of ECG signal based on Hadoop Framework"

[4] G.Vaishali1, V.Kalaivani2 "bigdata analysis for heart disease detection system using mapreduce technique"

[5] Cheryl Ann Alexander and Lidong Wang "Big Data Analytics in Heart Attack Prediction" DOI: 10.4172/2167-1168.1000393

[6] Hsieh JC, Li AH, Yang CC (2013) Mobile, cloud, and big data computing: Contributions, challenges and new directions in telecardiology. Int J Environ Res Public Health 10: 6131-6153

Asiya M. Al-Busaidi and Lazhar Khriji "Digitally Filtered ECG Signal Using Low-Cost Microcontroller"