

PULSE: A Health Monitoring Solution for Astronauts Utilizing Big Data and Machine Learning Methodologies

Swathi.K.A¹, Shashikala²P.G. Student (MTech), Department of Computer Science & Engineering, BNMIT, Bangalore, Karnataka, India¹Associate Professor, Department of Computer Science & Engineering, BNMIT, Bangalore, Karnataka, India²

ABSTRACT: We have seen a huge advancement in the last decade, in the amount of data we generate and collect and also in the ability to use technology to analyze and understand it. Naturally a lot of use cases in healthcare domain seems to be well suited for big data solutions. Healthcare research giants and private sector companies are investing heavily on making use of big data technologies to solve problems in healthcare. Big data solutions are widely in use to cure diseases, predict epidemics and improve the quality of life in general. Applying these solutions in the mission critical operations like space missions can be effective in maintaining the huge amounts of data being generated in real time.

Having understood the impact of big data on healthcare domain and the importance of astronaut's health in space missions, this paper proposes a way to monitor health of astronauts in long duration space missions. We plan to make use of big data technologies to run data through advanced models created by using machine learning technologies on the historical data collected during previous space missions.

This paper describes PULSE as an application that aids in monitoring the health of an astronaut by collecting a diverse range of health data, during various phases of space missions. It is evident that self-diagnosis coupled with identification of proper preventive responses on negative trends of astronaut's health are critical in order to keep astronauts health in safer range. By analysing the historical health data (Astronaut health record) for negative health trends, we can predict the possibility of abnormal health that might occur in near future and avoid the adverse consequences that can be brought into the space missions. Long-duration missions bring numerous risks that must be understood and mitigated in order to keep astronauts healthy, rather than treat a diagnosed health disorder.

KEYWORDS: PHM, Apache Spark, Logistic regression, MLlib, Big Data, Machine learning.

I. INTRODUCTION

Technology powered by human intellectual is a boon for healthcare. Providing quicker health interventions using technology is soon going to become the norm. Healthcare analytics are in great zeal which represents the systematic usage of various health data and related aspects developed by the application of analytical models, which includes quantitative, statistical, cognitive, contextual, predictive models, to drive decision making for management, planning, learning and measurement in healthcare, that are based on facts.

In this fast paced world, attributes like aging population, lifestyle changes and lack of proper diet, causes continued pressures on healthcare operations around the globe. These trends along with the digitization of patient and health data through improvements in information technology, leads to the generation of large quantities of primary and secondary data in the healthcare domain [1].

Space is the dimension in which humans have enormous opportunities to explore and expand their reign into new territories. Preparations for inter-planetary missions, either robotic or crewed, add to this aspiration. Such missions, by

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 5, May 2017

mobilizing a substantial collective effort, constitute one of the main areas of development of space activities in the decades to come. In this context, the objective of the paper revolves around 3 significant aspects:

- Healthcare Analytics with respect to Long-duration Space mission.
- Utilization of Big data solution and Machine learning to support Prognostics and monitoring Astronaut's health.
- Overview of PULSE Astronaut PHM Analytics framework: A mixed mode operation of big data and machine learning driven health analytics solutions.

PULSE analytics framework along with related computing technologies can process huge amounts of real-time data to obtain tremendous insights for effective clinical decision making in the research of long-duration space mission and significantly helps in monitoring astronaut health. In this paper, there is an illustration and implementation of a healthcare analytics framework in association with Apache Spark for real-time healthcare monitoring and related analytics by integrating heterogeneous data sources like health monitors or sensors, and astronaut health datasets. One of the ultimate goals of this work is to provide an effective, scalable and secure framework for increasingly important astronaut health surveillance.

II. RELATED WORK

The main objectives of the paper include:

- 1. Application of Big data processing and Machine learning adaptation in Healthcare domain:**
Using new technology to monitor the health of astronauts and provide assistance during any negative impact. This involves huge amount of data collection and processing. To maintain huge amount of data like historical, behavioral, operational, environmental data we are using big data framework[2]. Acquiring information about the clinical data which is kind of unstructured data at every second should be very fast which includes processing and storing of unstructured data. Big data platform provides a solution for handling such aspects.
- 2. Visualizing prognostics and health of astronauts:**
Prognostic is an engineering discipline that is used to calculate and estimate any negative trends and fixing it in an efficient amount of time before it goes to major damage. Health of astronauts is very critical. So using prognostics technique will make the work easy and quick.
- 3. Recognizing health pattern through Live monitoring system:**
Clinical data will vary from one environment to another, so there is a need to handle and monitor them which can be effectively done through live monitoring system.
- 4. Detecting health problems in prior using assistive technology:**
If any negative health trends are detected in the clinical data, we can analyze it and suggest the required solution before it can cause major damages to the health of astronauts.

There are basically 2 control centres for monitoring the Long-duration space mission:

- 1) Command center:** It is an important component of a comprehensive emergency space exploration management solution. The centre services as a primary hub of all communications between spaceship and earth based monitors, with the capability of receiving and tracking status and directing response activities. As such, communications technologies including voice (telephony and radio), video and data are all essential to the operation of these centres. It is a facility that manages space flights, usually from the point of launch until landing or the end of the mission. It is part of the ground segment of spacecraft operations. A staff of flight controllers and other support personnel monitor all aspects of the mission using telemetry, and send commands to the vehicle using ground stations.
- 2) Spaceship center:** It controls all the activities within the spaceship and monitors the assignments of astronauts and communications from outside the earth's atmosphere.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 5, May 2017

There are 2 different kinds of data that needs to be assessed as part of this work:

- 1) **Environmental Data:** They are the data which is based on the measurement of environmental factors, the state of the environment and the impacts on ecosystems. We have considered Atmospheric temperature, Humidity, Spaceship oxygen level as the environmental parameters for our analysis.
- 2) **Astronaut's Health Data:** They are the data of astronaut's health that are continuously monitored in a space mission. We have considered Blood Pressure, Respiration rate, Blood Glucose, Body temperature, Body Oxygen level, Activity count, Body fat, Heart rate, Physical strength as health parameters for our analysis.

This work makes use of potential benefits of big data to monitor and assist the astronaut. Monitoring and assisting requires historical data as well as behavioral data that can be efficiently managed using big data technique. To increase the quality of spaceflight, it is necessary to monitor as well as to protect the health of astronaut. The real challenge is to provide a way to monitor and analyze the central place of information & visualization to his/her physical or any internal body condition in real time.

Typically an advanced vest is worn by an astronaut during such missions. These vests help in gathering health data of astronaut like pulse rate, heart beat etc. The environment data is collected through high performance sensors. By utilizing these data, PULSE Astronaut PHM Analytics framework will help analyze the health conditions of astronaut, identify negative health patterns and rectify the reasons for these negative health patterns and thereby provides qualitative performance in a space mission. The basic workflow of PULSE is as shown in Fig. 1.

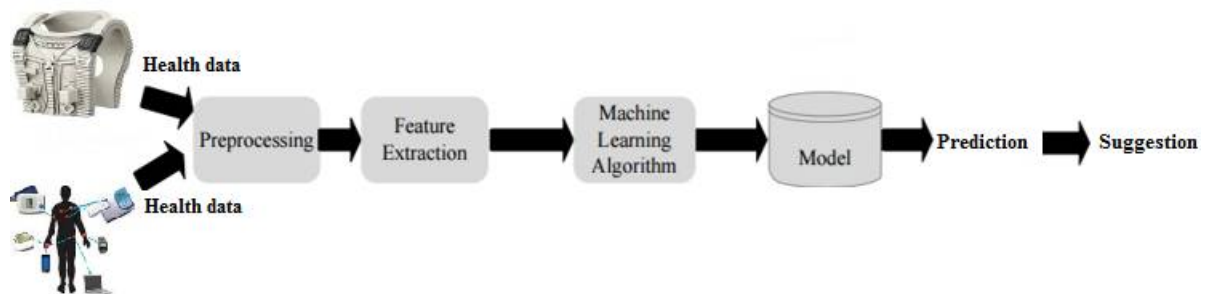


Fig.1. PULSE workflow

Apache Spark is one of the big data framework, which supports rapid development. Spark streaming allows to reuse the same code for batch processing, join streams against historical data, or run ad-hoc queries on stream state [3]. Build powerful interactive applications, not just analytics. Spark is mainly used for huge-scale streaming data processing. Spark has support to utilizeMLlib for machine learning, Graphprogramming, and Spark Streaming [6]. Spark is used as a core processing engine in this paper, which processes real time data collected from the device. Monitoring health data involves recognizing the pattern based on analyzing health and environment data.

Processing will be performed by Spark, recognizing the trend will be supported by machine learning algorithm and it also provides suggestions towards improvising astronaut's health. Spark has supportive libraries for machine learning; [4] ML has different sets of algorithms. Logistic regression algorithm is a classification algorithm, which evaluates the relationship with categorical dependent variable and one or more independent variable by estimating probabilities using a logistic function, which is the cumulative logistic distribution. Real time data is processed and it is possible to predict the trend by the model based on ML algorithm. Whenever negative trends appears, it will give an alert along with suggestion. This will help in monitoring health of astronauts and in providing health assistance to an astronaut.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 5, May 2017

III.SYSTEM DESIGN

A. SYSTEM ARCHITECTURE:

PULSE system has 4 main components, as represented in Fig. 2:

- **Data generator:** this phase contain health data and environment data in JSON format. This phase collect data from the sensor continuously.
- **Core Processing:** This module comprises of REST APIs which connects disconnected system, Spark is a platform for analysing real time data and identifying prediction using ML algorithm. ML algorithm in addition with training data builds a model for prediction. Knowledge extraction will discover the health trend of dataset depending on which, suitable health suggestions can be given to improve health conditions of the astronauts. Processed dataset is stored in FSDB for further usage.
- **Data storage:** It provides persistent storage to the processed historical data.
- **Data representation:** This component will display the analysed health values continuously and also produces graphical picture for the prediction.

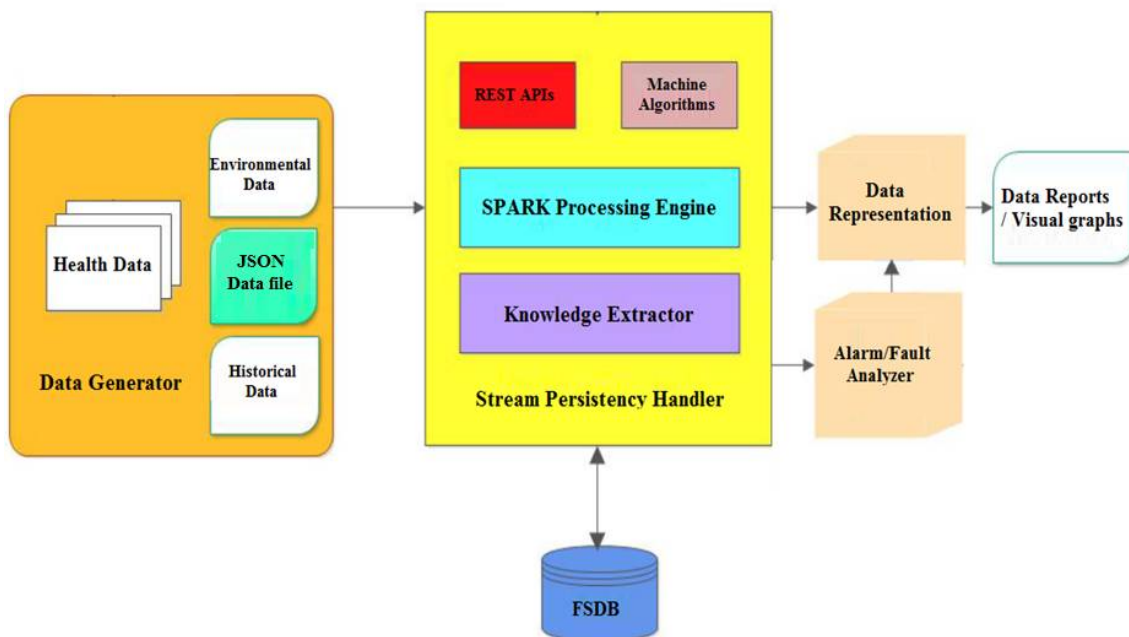


Fig.2. System architecture

The Spark processing engine associated with 3 different sub-systems processes historical data/training data, which will be used during one time model training. FSDB/HDFS store is used for persistent storage of analysed data. NodeJS/REST APIs, for developing server-side web application cross-platform runtime environment is useful; NodeJS/REST APIs acts as connectors between dashboard/visual graphs and the FSDB/HDFS.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 5, May 2017

B. ALGORITHM:

In our paper, we are making use of a data generator algorithm which can simulate the data generation as on a typical space mission. Focus was given to application of big data rather than data collection itself. Also historical data collected from previous space machines was gathered.

This paper proposes to use Apache Spark's MLlib to apply machine learning algorithm on the historical data gathered. Investigation went into choosing the right algorithm and the options evaluated included linear regression, Logistic regression, and Multi-class Logistic regression, linear and Gaussian kernels [5]. **Multi-class classification using sigmoid function** yielded the best results and this is what has been employed to predict/classify. This algorithm will also help in creating model and analyse real time health data pattern. If there is any diversion in pattern, it will provide alert with suggestions.

procedure intelligentHealthComputeEngine :

begin trainingPhase :

//load data set - training data

JavaRDD< LabeledPoint >[] splits = data.randomSplit(new double[] {0.6,0.4}, IIL);

//select hypothesis function - This is a sigmoid function

$h\theta(x) = g(\theta^T x)$;

$z = \theta^T x$

$g(z) = 1 / (1 + e^{-z})$

//Load the cost function - also called error function

Cost($h\theta(x), y$) = $-y \log(h\theta(x)) - (1-y) \log(1-h\theta(x))$ //where $h\theta(x)$ is hypothesis function as described before

//Apply gradient descent to find optimal values of theta which minimizes the cost function

Repeat{

$\theta_j = \theta_j - \alpha \partial / \partial \theta_j (J(\theta))$

}

//save the model into DB

model.save (sc, "target/temp/healthModel");

end trainingPhase

begin predictionPhase : //PREDICTION PHASE FOR STREAMING DATA

//Optimum theta values can predict result for streamed data by applying hypothesis function

//for a given matrix X what is prediction y

//where θ is our model, X is out input parameters and $h\theta(x)$ is our prediction

$h\theta(x) = g(\theta^T x)$;

end predictionPhase

We first divide the historical data set into 2 categories:

- 1) Training set
- 2) Test set in 60-40 ratio.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 5, May 2017

Feature scaling is not considered as the parameters seems to be in same range. The hypothesis function is given by - $g(z) = 1/(1+e^{-z})$, the cost function is of this form - $Cost(h\theta(x), y) = -y \log(h\theta(x)) - (1-y) \log(1-h\theta(x))$ and there are 3 classifiers for the below classes

1. Health is Normal
2. Health is deteriorating
3. Health is Abnormal

Gradient descent is applied to calculate the optimal θ parameters for each of the classifier.

```
Repeat {
 $\theta_j = \theta_j - \alpha \partial/\partial \theta_j(J(\theta))$ 
}
```

The model thus calculated is stored in FSDB which can be loaded for predicting/classifying the incoming data. The data generated in real time can be further used to strengthen the model.

Prediction phase includes choosing the classifier that yields the best result by applying $h\theta(x) = g(\theta^T x)$ and the data is classified into that category.

C. DATA FLOW DIAGRAM:

Historical data consisting of astronaut's previous space missions is fed into the spark processing engine. The spark engine calculates a model using Multi class logistic regression algorithm using MLlib. The Model is then saved to DB and loaded during prediction phase.

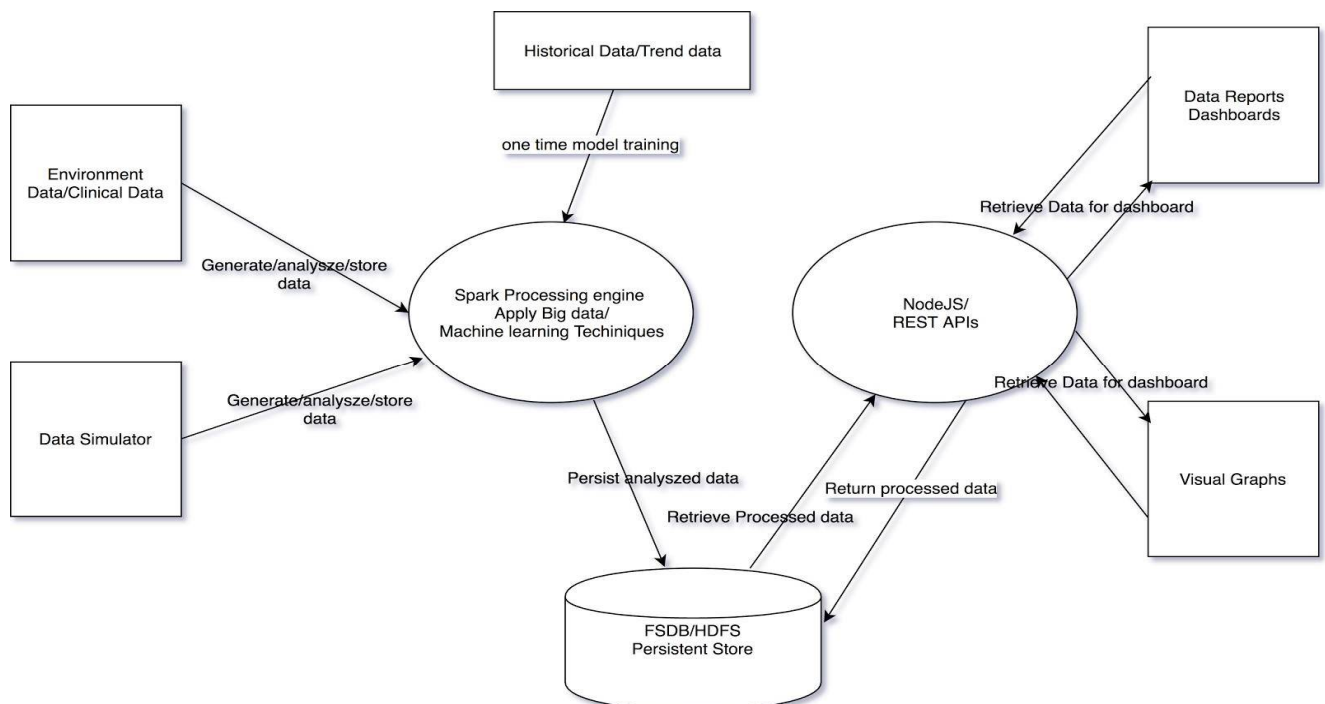


Fig.3. Data flow diagram

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 5, May 2017

The data flow of PULSE application is as represented in Fig. 3. Data representing astronaut health and surrounding environment is streamed through advanced sensors and data simulators and provided as input to the spark processing engine. Spark engine applies the model onto the data and predicts the class to which the data belongs (Health Normal, Health deteriorating, health Abnormal). The result is then persisted into the DB. In parallel, the frontend will periodically call the REST server which in turn queries the DB for prediction results. The response thus obtained will be presented in form of visually rich dashboards and graphs.

IV. EXPERIMENTAL RESULTS

PULSE follows a process which is instituted to record health problems and the onset of critical events, while health experts are learning about the health patterns progressively by analyzing the astronaut health record that gets generated continuously. This data can be used as input to a Pie-Bar Combination chart metrics.

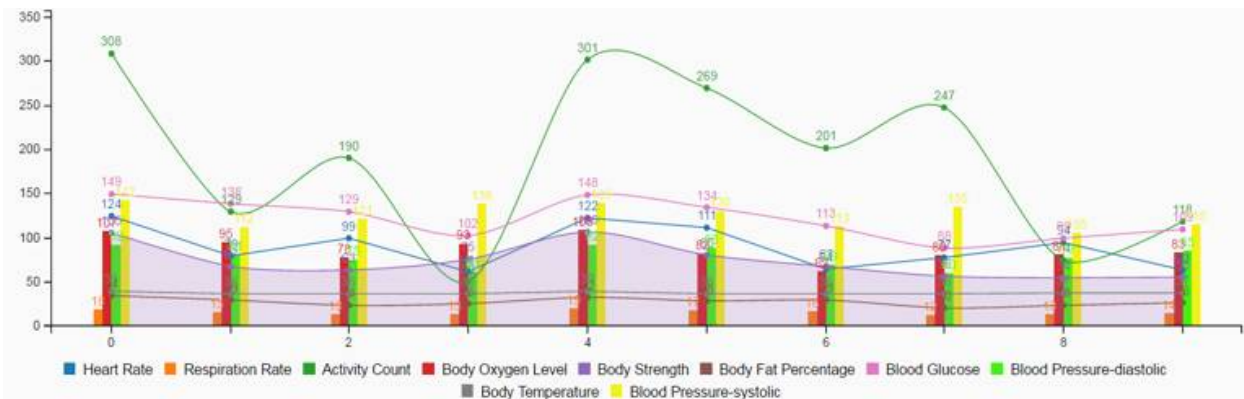


Fig.4. Graphical simulation of Health data set

This chart is used to highlight the vital health attributes. An example of a Pie-Bar Combination chart is shown in Fig. 4. It displays 12 health parameters overlaid with a cumulative frequency distribution. The y-axis plots the value of each health attribute. The x-axis plots the time interval (1 second per unit) against which health data is getting generated. Fig. 5 shows the results of health data generation by using an exemplar based precision modeling, capturing different health values.

```

2 1:123 2:147 3:96 4:17 5:139 6:311 7:39 8:33 9:107 10:109 11:60 12:109 13:62
2 1:100 2:113 3:67 4:13 5:124 6:218 7:36 8:20 9:73 10:86 11:9 12:65 13:59
2 1:146 2:162 3:118 4:22 5:159 6:383 7:40 8:39 9:111 10:129 11:76 12:128 13:88
2 1:138 2:162 3:101 4:24 5:152 6:390 7:44 8:38 9:136 10:145 11:96 12:135 13:82
1 1:117 2:139 3:84 4:13 5:110 6:192 7:37 8:20 9:51 10:67 11:53 12:65 13:36
0 1:84 2:101 3:73 4:17 5:131 6:189 7:36 8:22 9:62 10:72 11:-52 12:95 13:26
0 1:72 2:139 3:61 4:14 5:107 6:162 7:36 8:27 9:55 10:83 11:59 12:90 13:35
2 1:116 2:106 3:60 4:14 5:84 6:267 7:37 8:23 9:79 10:84 11:51 12:63 13:45
2 1:118 2:117 3:72 4:13 5:79 6:141 7:36 8:20 9:87 10:72 11:-29 12:43 13:50
2 1:142 2:157 3:109 4:22 5:164 6:382 7:44 8:43 9:145 10:133 11:86 12:137 13:83
0 1:89 2:130 3:61 4:12 5:81 6:136 7:37 8:24 9:79 10:83 11:-26 12:99 13:29
0 1:63 2:116 3:89 4:17 5:92 6:234 7:37 8:27 9:67 10:92 11:-37 12:80 13:39
0 1:91 2:135 3:89 4:13 5:83 6:287 7:36 8:23 9:80 10:63 11:12 12:91 13:59
2 1:115 2:135 3:75 4:12 5:105 6:142 7:36 8:29 9:82 10:86 11:31 12:40 13:41
2 1:147 2:164 3:114 4:22 5:167 6:385 7:40 8:39 9:112 10:135 11:97 12:138 13:76

```

Fig.5. Health and Environmental data generation in real-time

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 5, May 2017

Fig. 6 indicates the graphical representation of different classes of health data outcomes where “Health normal” class is recognized as “Success alarms”, “Health deteriorating” class is recognized as “Warning alarms” and “Health abnormal” class is recognized as “Danger alarms”)

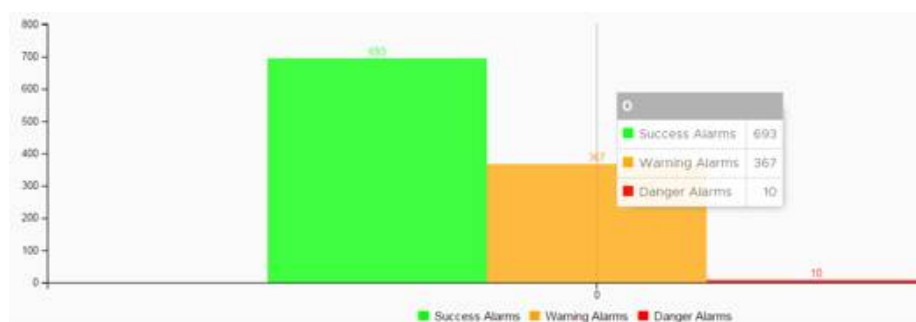


Fig.6. Health Event alarms

V.CONCLUSION

The intersection of Big data solutions for adapting huge amounts of health data combined with Machine learning algorithms to predict the trends of health data patterns, will significantly enhance quality of data streaming and its analysis. This will create the most efficient and productive results in the long duration space missions. It may be long or short duration spaceflight, astronaut’s health is continuously facing several health risks, due to hostile environment. Health monitoring of astronaut is needed, if any diversion in health trend has to be notified and alerts must be provided on the health status in order to provide excellent health support. Monitoring the health of the astronaut involves huge amount of health and environment data set, and need to handle streaming data, so speed of data collection and response should be very fast. Big data technology and Machine learning techniques forms a mixed mode platform for storing and processing huge amount of data while handling variety of data with high velocity, while determining the effects on future health events. PULSE is an idea which focuses on creating such an adaptive infrastructure to achieve these goals efficiently.

REFERENCES

- [1] Yu-Chong Tai, “Improving Healthcare Provider Performance with Big Data”, IEEE Transactions on healthcare support, Vol 2nd, 2014.
- [2] Ramkrushna C. Maheshwar “Survey on high performance analytics of bigdata with apache spark”, IEEE Transactions on Big data framework, Vol 2, no .9, 978-1-4673-9545-8, 2017.
- [3] Jian Fu, Junwei Sun , “SPARK – A Big Data Processing Platform for Machine Learning”, IEEE Transaction On intelligent technology, Vol. 3, no. 978-1-5090-3575-5, 2016
- [4] S. B. Kotsiantis · I. D. Zaharakis · P. E. Pintelas, “Machine learning: a review of classification and combining techniques” © Springer Science+Business Media B.V. Vol. 1, May 2007
- [5] Catherine Combes, Farid Kadri, Sond'es Chaabane, “PREDICTING HOSPITAL LENGTH OF STAY USING REGRESSION MODELS: APPLICATION TO EMERGENCY DE-PARTMENT”, Conference Francophone, Optimisation of Simulation, Nov 2014.
- [6] Satish Gopalani, “Comparing Apache Spark and Map Reduce with Performance Analysis using K-Means”, International Journal of Computer Applications, Volume 113 – No. 1, March 2015