TEST Engineering & Management

January - February 2020
ISSN: 0193 - 4120 Page No. 14009 - 14014

Integration of Hadoop and IOT for better analytics

.Manthru Naik[1], Dr.K.UmaPavanKumar[2],Dr.M.Praveen Kumar[3]
[1] Assistant Professor,[2]Associate Professor, [3]Professor
[1,2,3]Department of CSE, Malla Reddy Institute of Technology Hyderabad, India
[1] manthru440@gmail.com, [2] dr.kethavarapui@pec.edu,[3]praveen.hawassa@gmail.com

Article Info
Volume 82
Page Number: 14009 - 14014
Publication Issue:
January-February 2020

Article History
Article Received*: 18 May 2019*
Revised*: 14 July 2019*
Accepted*: 22 December 2019*
Publication*: 26 February 2020*

Abstract

The new sample inside the exam and normal applications is internet of factors (IOT). ThefunctionalbenefitsofIOTarerangingfromsmarthousetosmartcities.Themainpurpose of IOT is to contain specific gadgets intelligently and talking among the gadgets without human mediation. The prevailing exchange predominantly centers round utilising the restriction of exam in IOT and resolves the functionality troubles of the mass information created byIOT.
The proposed concept offers the usage of Hadoop diploma to store the facts and from that facts performing exam for higher usage of IOTcommunications.
The importance is clarified with a few consistent conditions in which there's right mixture of Hadoop degree and IOT. To preserve the extremely good education of the facts Hadoop allotted document gadget (HDFS) may be applied, and to ingest the information from outer stages we can utilize Sqoop or Flume.
The facts available in HDF test be used to system with the use of Map reduce(MR) method. When the facts is available in HDFS the exam may be completed with Hive, Pig or R on the subject of device getting to know or information miningtechniques.
The result of the proposed belief is coordination of Hadoop and IOT tiers with a added collectively casing work which fits the mixture of Hadoop and IOT, stockpiling arrangements to cope with mass statistics, getting ready of the placed away facts and utilising investigation if you need to correctly serve considered one of a kind companions.

Index Terms: *Hadoop, IOT, Analytics, Storage, Map Reduce*

INTRODUCTION

ML Algorithms inIOT

Machinelearning is key component in the processing of data with the major entity Table which includes row along with column family, the column qualifier which specifies the data which is availableinthatrow.Foreachandeverynewentrythetim estampalongwithvaluewillbeadded. ML includes the algorithms in the flavors of supervised learning, unsupervised algorithms and reinforcementalgorithm.

Supervised learning algorithms having the input data with labels, and based on the learning experience when ever the new data is given it produces the correct outcome. Classification, regression are the types of the supervised learning.
UnsupervisedLearningalgorithmshavingtheinputwith outanylabelsandallowingthealgorithm withoutanyguidance.ClusteringandAssociationarethe mostcommonlyusedalgorithmsinthis category.
Reinforcementlearningisdynamicdecisionmakingkin dofthealgorithmwheretheprocesstakes the decision so as to maximize the performance of the task. The best possible bahavior can be observed and the process takes care ofit.

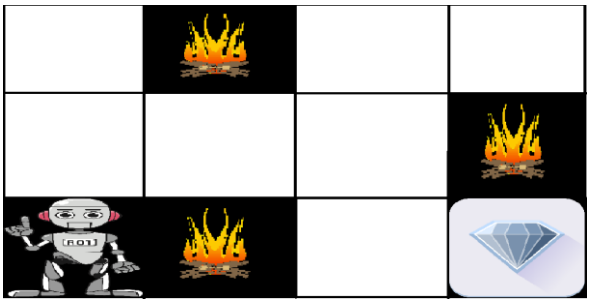

Figure: best possible way by the Robot to reach the diamond

Published by: The Mattingley Publishing Co., Inc.

TEST Engineering & Management

January - February 2020
ISSN: 0193 - 4120 Page No. 14009 - 14014

The reinforcement learning in the above problem explains us like the reachability of the robot to the diamond by avoiding the fire in the path.

IOT Applications

IOT nowadays playing a vital role starting from the range of smart home to smart city, thefollowing figure shows the application areas of IOT in various fields.

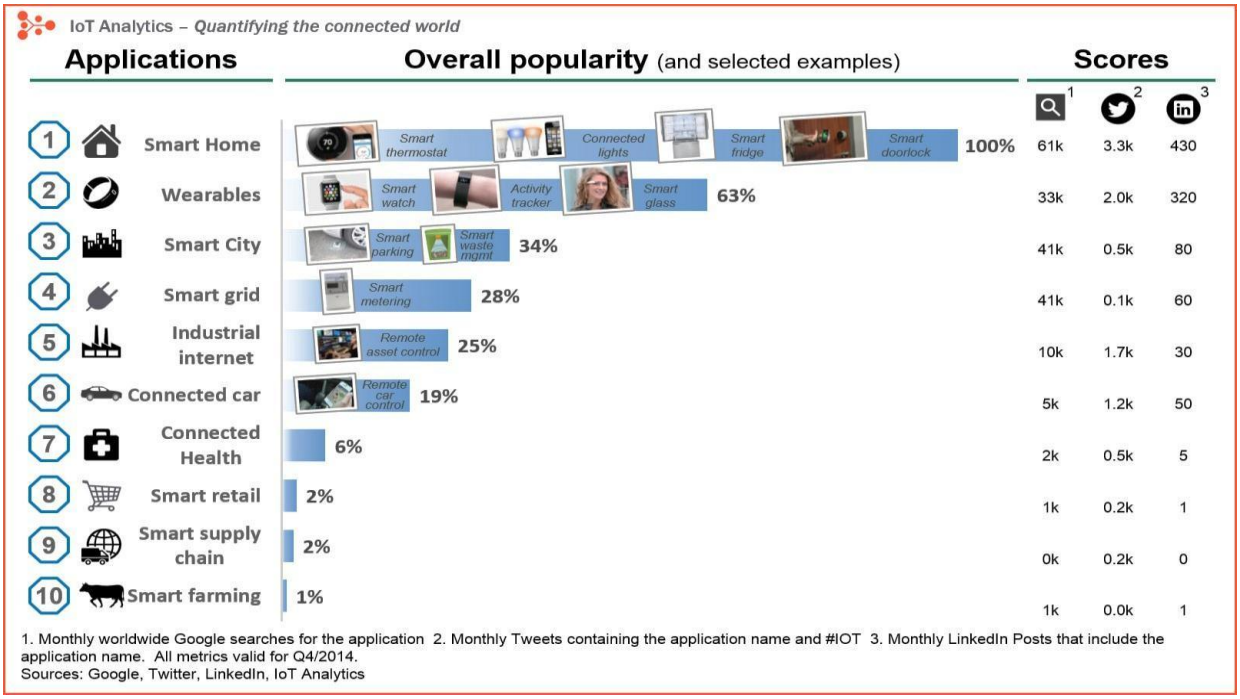

Figure:The scores of various IOT based Applications.

IOT and Analytics

Analytics mainly helps us to predict the future possibilities based on the given input data. There are many algorithms the most commonly used algorithms are as follows:

1. Regression
2. Clustering
3. Decision Trees
4. Random Forest
5. Classification
6. Descriptive statistics
7. Prescrptive Statistics
8. Predictive Statistics

Example:

```
##Cluster Analysis
setwd("C:/Users/haarshadatta/Desktop/uma")
library(arules)
data<- read.csv("groceries.csv",sep=",") data
View(data)
data<-read.transactions("groceries.csv",sep=",")
data
summary(data) inspect(data) inspect(data[1:3])
inspect(data[1001:1003]) itemFrequency(data[,1])
itemFrequency(data[1:6]) #more than 0.08 support
itemFrequencyPlot(data,support=0.08) #top 5 items
itemFrequencyPlot(data,topN=5)
#top 25 items itemFrequencyPlot(data,topN=25)
#applying Model
m1<-apriori(data) m1
summary(m1)
m2<-apriori(data,parameter =
list(support=0.007,confidence=0.5)) m2
```


Published by: The Mattingley Publishing Co., Inc.

January - February 2020
ISSN: 0193 - 4120 Page No. 14009 - 14014

summary(m2) inspect(m2[1:3]) #sorting rules by Lift

inspect(sort(m2,by="lift")[1:5])

inspect(sort(m2,by="lift")[1:15]) #specific Rules

m3<-apriori(data,parameter = list(support=0.007,confidence=0.2),appearance=list(default="rhs",lhs=c("whole milk","soda")))

summary(m3) inspect(m3[1:3])

Result:

```
[1] {citrusfruit,rootvegetables}                => {other vegetables} 0.0103
711240.5862069   3.029608102
[2] {root vegetables,tropicalfruit,whole milk} => {other vegetables} 0.0070
157600.5847458   3.022057  69
[3] {rootvegetables,tropicalfruit}              => {other vegetables} 0.0123
029990.5845411   3.020999121
[4] {tropicalfruit,whipped/sourcream}           => {other vegetables} 0.0078
291810.5661765   2.926088  77
[5] {rootvegetables,wholemilk,yogurt}           => {other vegetables} 0.0078
291810.5384615   2.782853  77
>inspect(sort(m2,by="lift")[1:15])
     lhs                                            rhs                suppo
 rt    confidencelift      count
[1]  {citrus fruit,root vegetables}              => {other vegetables} 0.010
3711240.5862069   3.029608 102
[2]  {root vegetables,tropicalfruit,whole milk}  => {other vegetables} 0.007
0157600.5847458   3.022057  69
[3]  {root vegetables,tropical fruit}            => {other vegetables} 0.012
3029990.5845411   3.020999121
[4]  {tropical fruit,whipped/sour cream}         => {other vegetables} 0.007
8291810.5661765   2.926088  77
[5]  {root vegetables,wholemilk,yogurt}          => {other vegetables} 0.007
8291810.5384615   2.782853  77
[6]  {pip fruit,root vegetables}                 => {other vegetables} 0.008
1342150.5228758   2.702304  80
[7]  {pork,root vegetables}                      => {other vegetables} 0.007
0157600.5149254   2.661214  69
[8]  {domestic eggs,root vegetables}             => {other vegetables} 0.007
3207930.5106383   2.639058  72
[9]  {tropical fruit,wholemilk,yogurt}           => {other vegetables} 0.007
6258260.5033557   2.601421  75
[10] {rolls/buns,root vegetables}                => {other vegetables} 0.012
2013220.5020921   2.594890 120
[11] {root vegetables,whipped/sour cream}        => {other vegetables} 0.008
5409250.5000000   2.584078  84
[12] {root vegetables,yogurt}                    => {other vegetables} 0.012
9130660.5000000   2.584078127
[13] {butter,yogurt}                             => {whole milk}       0.009
3543470.6388889   2.500387  92
[14] {butter,root vegetables}                    => {whole milk}       0.008
2358920.6377953   2.496107  81
[15] {other vegetables,tropicalfruit,yogurt}     => {whole milk}       0.007
6258260.6198347   2.425816  75
```


Published by: The Mattingley Publishing Co., Inc.

January - February 2020
ISSN: 0193 - 4120 Page No. 14009 - 14014

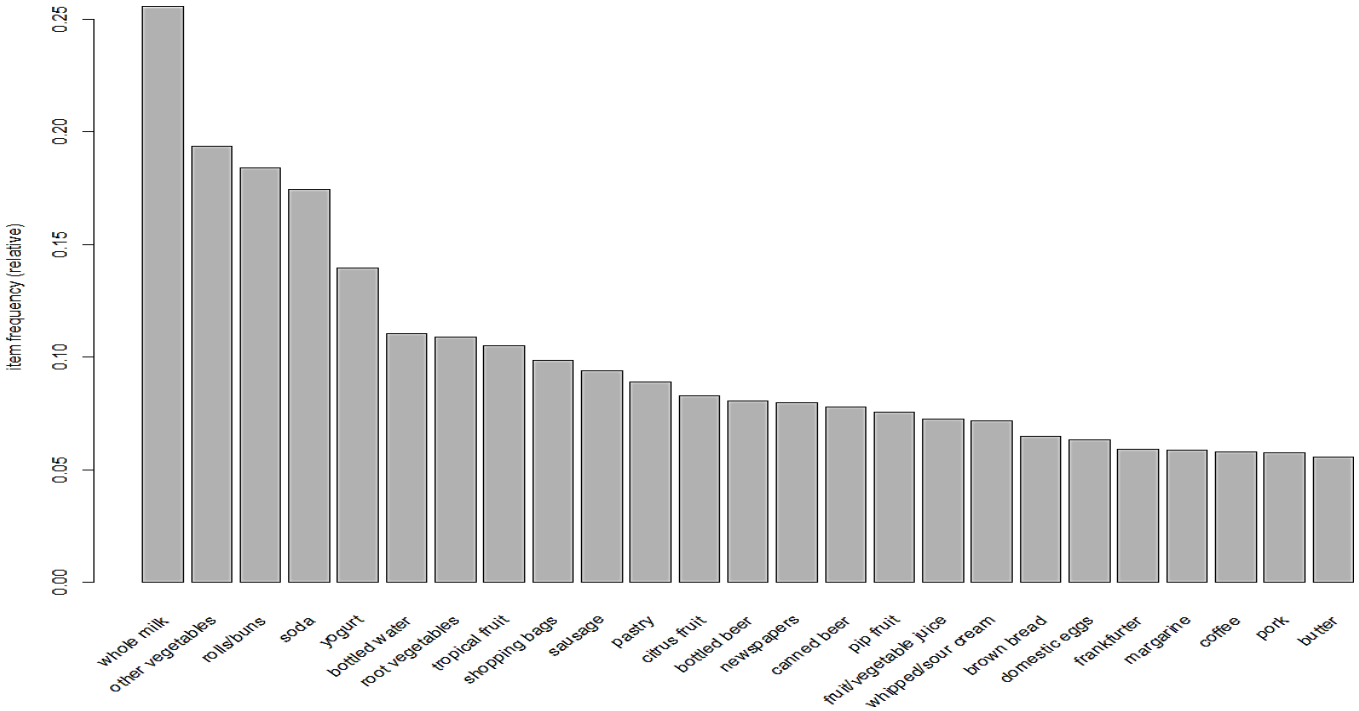

Figure: Top 25 items identification with Association

Incaseofagriculturemoistureobservation,humidityand otherparametershelpsustogivesome suggestions about the crops to the farmers. The same association rule can be applied to the input data, the rules which influence the crop condition and information about weather forecasting based on the past data like rain fall data, temperature, humidity andothers.

The following is the example prediction so as to explain about the temperature data in Andhra Pradesh area.

Maximum Temperature: Daytime maximum temperatures will be in the range of 38- 44C.

Rainfall: Isolated thundershowers will occur at a few places over AP and Telangana

Discomfort Index (DI): Daytime Discomfort Index will increase, to be in the range of 29 to 32.

Minimum Temperature: Night time temperatures will be in the range of 29-34C.

Humidity: Humidity will be in the range of 10-25%.

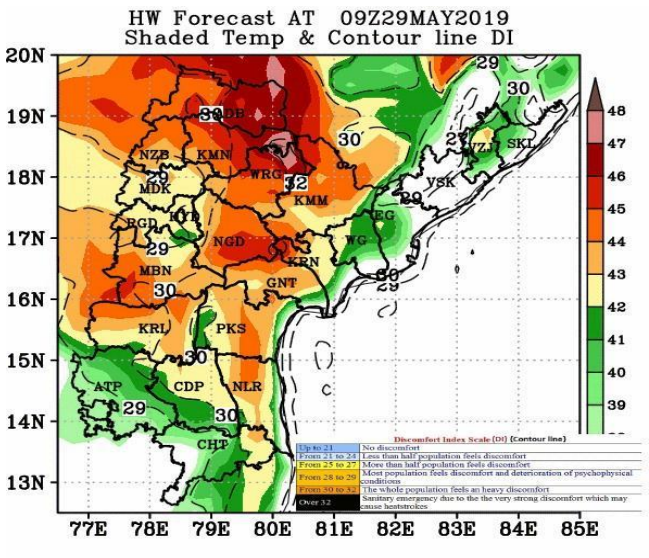

Figure:Temperature Prediction

The devices helps us to find out various measures and the applicability of the analytics with the helpofRprogramminglanguage/Python.RODBCisone ofthepackagebasicallyusedtohandle embed the code with IOT devices. The IOT programming still complex and R/Python/Java are providing few techniques or packages so as to construct IOT basedlogics.

Published by: The Mattingley Publishing Co., Inc.

TEST Engineering & Management

January - February 2020
ISSN: 0193 - 4120 Page No. 14009 - 14014

SomeofthedeviceswhicharemostcommonlyusedIOT devicesintheagriculturearedisplayed here.

Figure : IOT device used in Agriculture.

Research Issues and future scope of IOT and Analytics

There are many research gaps and issues were there in the integration of IOT and Hadoop up to theadvancementofAnalytics.Herewareprovidingafe wpointsasaresearchissuesandfuture scope of the above heterogeneoustechniques.

- Universal framework to store and process the IOT data with the Help of HDFS andMR
- Common procedure to embed the IOT, bigdata andAnalytics
- A file system to handle IOT sensor data, HDFS and Analyticaldata
- A simplified MR model to process the IOT data while processing the unstructured data.
- A common packages need to design in R/Python so as to handle IOT based data processing.
- A simple procedure to establish a communication among the IOT, Hadoop andAnalytics code with reasonable time and spacecomplexity.
- The biggest challenge is to handle the inter process communication between the IOT sensor data, HDFS and the implementationtechnology.

CONCLUSION

The chapter explained the basics of Hadoop and Hadoop eco system with all the possible applications that can be implemented with various eco system tools. The IOT fundamentals, IOT integration with Hadoop eco system along with ML support has been explained.

The usages of ML with various possible algorithms were described so as to work with various formats of the data in the implementation of the algorithm. The article explained the IOT devices used in the field of agriculture and have shown the temperature graph with the predictions of the future temperatures. The article also explained various research issues and future directions possible in the IOT based analytics in the context of the big data with Hadoop.

TEST Engineering & Management

January - February 2020
ISSN: 0193 - 4120 Page No. 14009 - 14014

REFERENCES

[1] S. Lohr, "The period of massive information," N. Y. Times, vol. Eleven, 2012.

[2] S. Goad, "From Databases to Big Data.," IEEE Internet Comput., vol. Sixteen, no. Three, 2012.

[3] P. Zikopoulos, C. Eaton, and others, Understanding big facts: Analytics for huge business magnificence hadoopand gushing records. McGraw-Hill Osborne Media,2011.

[4]A.McAfee, E.Brynjolfsson, T.H.Davenport,D.J.Patil and D.Barton,"Bigdata,"Manag.Revolut.Harv. Transport Rev, vol. Ninety, no. 10, pp. 61–sixty seven, 2012.

[5] R. Appuswamy, C. Gkantsidis, D. Narayanan, O. Hodson, and A. Rowstron, "Scale-up as opposed to Scale-out for Hadoop: Time to reexamine?," in Proceedings of the fourth every year Symposium on Cloud Computing, 2013,p. 20.

[6] A. S. Tanenbaum and M. Van Steen, Distributed frameworks. Prentice-Hall,2007

[7] C. P. Chen and C.- Y. Zhang, "Dataintensive programs, difficulties, systems and advancements: A review on Big Data," Inf. Sci., vol. 275, pp. 314–347,2014.

[8] T. B. Murdoch and A. S. Detsky, "The inescapable utilization of substantial records to medicinal offerings," Jama, vol. 309, no. Thirteen, pp. 1351–1352,2013.

[9] I. Mashal, O. Alsaryrah, and T.- Y. Chung, "Execution assessment of proposal calculations on Internet of Things administrations," Phys. Detail. Mech. Its Appl., vol. 451, pp. 646–656,2016.

[10] K. Shvachko, H. Kuang, S. Radia, and R. Chansler, "The hadoop conveyed report framework," in 2010 IEEE twenty sixth symposium on mass stockpiling frameworks and improvements (MSST), 2010, pp.1–10.

[11] J. Senior member and S. Ghemawat, "MapReduce: progressed information making ready on big corporations," Commun. ACM, vol. 51, no. 1, pp. 107–113, 2008.

[12] J.Y.Monteith, J.D.McGregor and J.E.Ingram,"Hadoop and its Evolving Ecosystem.,"inIWSECO@ ICSOB, 2013, pp.Fifty seven–68.

[13] K. Ting and J. J. Cecho, Apache Sqoop Cookbook. O'Reilly Media, Inc., 2013.

[14] S. Hoffman, Apache Flume: Distributed Log Collection for Hadoop. Packt Publishing Ltd, 2013.

[15] S. Haloi, Apache ZooKeeper Essentials. Packt Publishing Ltd, 2015.

[16] M.K.IslamandA.Srinivasan,ApacheOozie:The Work flow Scheduler for Hadoop.O'ReillyMedia,Inc., 2015.

[17] C. Olston, B. Reed, U. Srivastava, R. Kumar, and A. Tomkins, "Pig latin: a notso-unknown dialect for statistics processing,"inProceedingsofthe2008ACMSIGMODinternationalconferenceonManagementof data,2008, pp.1099–1110.

[18] H. Bansal, S. Mehrotra, and S. Chauhan, Apache Hive cookbook.Packt Publ.,2016.

[19] E.Alpaydin, Introduction to machine learning (adaptive computation and machine learning series).TheMIT Press Cambridge, 2004.

[20]. Uma Pavan Kumar Kethavarapu, "Different Computing models in Hadoop eco framework alongside the point of view of investigation utilising R and Machine mastering", Vol. 14 CIC 2016 Special Issue International Journal of Computer Science and Information Security (IJCSIS), PP-17-23.

[21]. C. P. Chen and C.- Y. Zhang, "Information escalated applications, problems, strategies and innovations: A look at on BigData", Inf. Sci., vol. 275, pp. 314347,2014.

Published by: The Mattingley Publishing Co., Inc.