

Analyzing Olympic History and Unstructured Sentiments with the help of Big data using MapReduce framework

Shareen Shahana Marottikal Khader

X18141404

MSC Data Analytics

National College of Ireland

Abstract— The Olympic Games are considered the world's foremost sports competition with more than 200 nations participating. After every four years, the Olympic Games are held, with the Winter and Summer Games alternating by occurring every four years but two years apart. This research paper analyses different participants from 1896 to 2016 through the Big data ecosystem. The Olympics medal-winning is a pride for every patriotic citizen of the country. MapReduce has been done to get the analysis of medals and trend of athletes throughout the world. This research, we analyze a structured Historical Olympic participation data and an unstructured Image with Survey of US citizens interest in watching the game to answer the business queries. The research uses Hadoops distributed computing capabilities for large datasets. The pre-processed structured datasets are stored in MySQL and the unstructured dataset is stored in HDFS as Input to Bigdata frameworks. The unstructured image data is processed with the help of OCR techniques. The Sqoop is used to import data from MySQL to HDFS. The hive, pig Java MapReduce is used to perform the Business Objective tasks. The Result of MapReduce is directly stored into an HBase which is a NoSQL database. The results from HBase tables are extracted using java in the form of CSV. The Results are visualized to know the solution to the business queries.

Keywords— MapReduce, Hadoop, Java, Sqoop, MySQL, Pig, HDFS, Hive, HBase

I. INTRODUCTION

The ancient Greeks have started gathering all the men after every four years for the sporting events in the town Olympia. The awareness of having modern Games was proposed in the mid-19th century. And became the world event in 1896. It was postponed because of wars; they have held since then after every four years around the world in different cities. Olympics or the modern Olympic Games or are foremost international sporting events including summer and winter sports competitions in which thousands of athletes from around the world take part in a variety of competitions.

A. Objective and Research Questions

The objective of the research is to find meaning full insight into the Olympic medal-winning. The research question is can we use MapReduce Bigdata framework to analyses the Olympics historical data with the help of unstructured sentiments to solve Business queries? The business queries are listed below.

- 1) Which are the top 10 countries with most Olympic gold medals?
- 2) Comparison of women Olympians who won the gold medal with men Yearly?
- 3) How are sentiments affecting the US Performance in Olympics?

- 4) Which game in the Olympics has the highest participation and which one has the lowest?
- 5) Is there any difference in season-wise participation in Olympics?
- 6) Which game demands the tall and which one demands thin people in the Olympics?
- 7) Top 10 gold medal hunters in Olympics history?
- 8) Most preferred cities hosting the Olympic game season-wise?

B. Motivation for Research

In 2008,2012 and 2016 Usain Bolt consecutively won Olympics 100m and 200m . and he is the only sprinter to do so. He won a total of nine Olympics gold medals. The Motivation of the research is identifying such facts and milestones so far. C. Thirumalai et al. (2017) found that united states as the highest possibility to the next Olympic due fact that the US is the country having the maximum number of medals. She used a machine-learning algorithm to prove the probability.

II. RELATED WORK

Olympic medals have consistently been engaged by the world's media and onlookers. The reason for this paper is to investigate the relationship of award count with the demographics, gender-wise division along with analysis of height and weight with any specific games. positioning dependent on gold, decorations and decoration score is progressively ideal to ground-breaking nations. Every nation has a quest for gold. And the countries performance has been judged by the number of golds won by the country.

The total number of medals won by the country, many researchers have done the have analyzed performance of the participating candidates. Nonparametric techniques are oboe of the popular method which has been used throughout. Along with this data envelopment analysis peer decision-making units. This is for making efficiency evaluation which makes assumptions on the production function on multiple inputs and outputs. (Li Liang Chen and Morita 2008)(Lins et al.) assumed the different nations- imposed AR constraints and applied context dependents. Advancement of the DEA has been proposed by (Wu Liang and Yang 2009). Heuristic search model optimized this model (Li chen, Liang & Xie 2012) are developed the measures and efficiency of each nation and it is two individual stages the same model has been introduced to analyses the medals in 2012 London summer Olympics. The study measures the performance of participating nations in a two-stage process. In prior two-

stage DEA methods, two common methods are used for aggregating individual stages process. The weight additive manner describes the efficiency of the entire two stages process as a weighted average of the efficiency of individual stages. (performance evaluation of participating) also worked on the London summer Olympics games with the DEA model and came out with the conclusion that there is a correlation between input participants with the medal. winner. Ey have used heuristic search technique for the assumption of the medal winner in the next game with respect to sex and demographic condition.

G. Ming et al. (2017) did research on technical and athletic level analyzes on track, field sports scores man the status and finally, the value for the country are calculated to plan a better strategy for Olympics game-planning china. The author found that by analyzing the historical data the probability of getting medals in long jump, javelin, 50km, 110m hurdle race, triple jump, 20 km walking race, marathon and 5000m are very high while china has low performance in shot put, 400m hurdle race, 800m, 1500m, and hammer, etc.

Binhui Wang et al. (2011) found that The powerful countries having a good ranking based on the gold medal the keep. The less populated islands countries having more medal per capita score. The medal score ranking based on GDP favorable on central Asia and poor Africa.

T. A. Ashwitha et al. (2017) used the Hadoop MapReduce framework to analyze large movie dataset. The authors also climes that

They notice a drastic improvement in the processing time when compared with traditional configuration of systems.

Overall games, competitors and mentors are progressively working with huge information and examination to press each and every understanding out of each drop of information accessible. The possibility of Olympic decorations and accomplishment on the worldwide stage mean at the tip-top level, there is an extreme strain to be on the bleeding edge of investigation. The thought is that by gathering all of the information about each competitor who enters the preparation program, new contestants can be coordinated against profiles of previous participants, to recognize the methodology well on the way to transform every person into a hero. The specialized aptitudes expected to construct a Big Data-driven examination framework are not local to the universe of first-class sports researchers and physiologists. So, the big data is used Training athletes, predicting the outcome and games winner.`

Gracenote has revealed a prediction before Olympics 2016 which country will get how many medals and who will definitely win in which section. The trainers from various countries take the help of big data to train their trainee. For the next season. Performance analysis with their body, height, BMI everything will be analyzed and according to that will give chart for training which will bring up the success rate.

III. METHODOLOGY

A. Dataset

The 120-year history Olympics is downloaded from Kaggle¹ as structured data. The dataset contains 271,116 rows and 15 columns. The above dataset data Contains the following columns

- 1) ID - Unique number for each athlete
- 2) Name - Athlete's name
- 3) Sex - M or F
- 4) Age – Integer
- 5) Height - In centimeters
- 6) Weight - In kilograms
- 7) Team - Team name
- 8) NOC - National Olympic Committee 3-letter code
- 9) Games - Year and season
- 10) Year – Integer
- 11) Season - Summer or Winter
- 12) City - Host city
- 13) Sport – Sport
- 14) Event – Event
- 15) Medal - Gold, Silver, Bronze, or NA

The Unstructured dataset used for the research is downloaded from Statista². It contains details of the survey contacted about America's interest in the Olympics. The data extracted from images are year percentage of people think a great deal and the percentage of people that thinks not much.

B. Data Pre-Processing

The Data pre-processing is done Using RStudio. The Structure data has to be cleaned in order to upload into the Big data framework. The Unstructured data has to be extracted into the text with the help of Optical character recognition libraries in R. The following are the data preprocessing done as part of the research.

- 1) *Unique ID Generated:* A unique identifier is generated to track each column due to lack of unique identifier in the dataset.
- 2) *NA Values are handled:* NA values are replaced with the median value of the data frame column. The replace columns are height, weight, age, medal, etc.
- 3) *Unwanted columns deleted:* There were few columns not used by the research. Hence unwanted columns are deleted. The columns deleted by the project are Team, Game, and Event.
- 4) *Extracting Text from Image:* The Unstructured data is processed and extracted the text using OCR technique. OCR library used in R is tesseract.it read the image into editable and searchable data.

Once the preprocessing is done the data is uploaded into MSQl as the Input for further processing. The following section will explain how the Big data use this data to process Business logic and find the solution.

¹<https://www.kaggle.com/heesoo37/120-years-of-olympic-history-athletes-and-results>

²https://infographic.statista.com/normal/chartoftheday_5523_americans_interest_in_olympics_at_historic_low_n.jpg

C. Approach

The Approach used for the research is discussed below.

1) System Configuration

Software	Version
Apache Hadoop	2.9.2
Apache Pig	0.17.0
Apache Hive	2.3.5
Apache Sqoop	1.4.7
Apache HBase	1.4.10
MySQL	5.7.27

The research is done on an Ubuntu 18.04 version is installed in virtual box with 8 GB RAM and 50GB hard disk. The Virtual box is configured with the following technologies

2) System Architecture

The figure below shows the architecture for the MapReduce framework in bigdata.

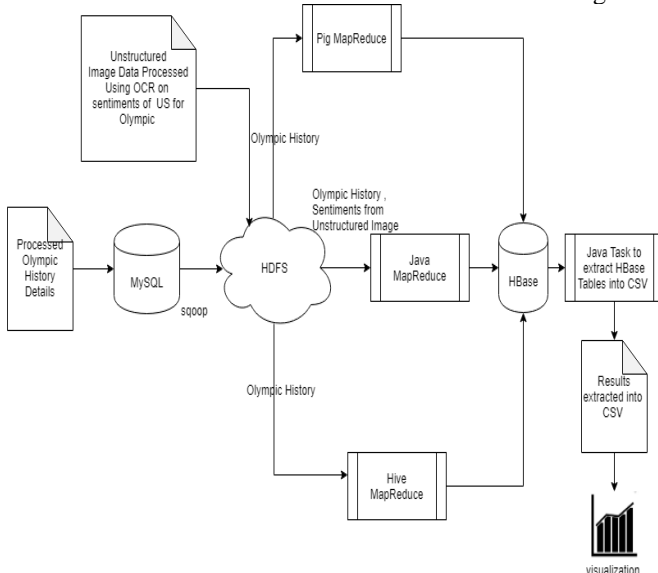


Figure 1: the architecture of the system.

The processed structured file is loaded into MySQL after creating the table. The loaded Historical data is imported into HDFS with the help of Sqoop. The unstructured Sentiments of Americans are uploaded into HDFS with the help of HDFS inbuilt commands. Pig, hive, java MapReduce are used to perform the tasks. These tasks are decided according to the Business Objectives. The research performs eight business queries. The map reduced outputs are uploaded into HBase directly from the respective scripts. Finally, the outputs are extracted into CSV from HBase Table with the help of java code. The solution for the business Objective is visualized with the help of Tableau or Power BI. **All the task discussed above is automated with the help of a shell script which includes downloading the data, preprocessing the data inserting into MySQL followed by HDFS with help of**

Sqoop. Then execution of pig, hive and java MapReduce, uploading data into HBase and extracting the HBase table into CSV.

a) MySQL

The MySQL Table is created which matches our historical data of Olympics. The data is loaded into MySQL with the help of following queries

[illegible]

Figure 2: MySQL File import

The total of 268410 records is loaded into the MySQL table. few records or failed to load MySQL due to invalid data format. The figure below shows the total record loaded in the DB.

```
mysql> select count(*) from olympics_History;
+-----+
| count(*) |
+-----+
|    268410 |
+-----+
1 row in set (0.09 sec)
```

Figure 3: total count of record in MySQL table

b) Sqoop

The Sqoop is a tool which used to transfer data from Hadoop distributed system to Relational Databases and vice versa. It considered being the fast and reliable tool for data transfer between Hadoop and MySQL here P. M. Bante et al. (2017). hence Sqoop is considered

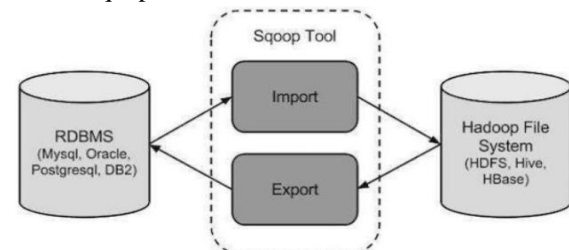


Figure 4: workflow of Sqoop.

The command given below is used to upload MySQL Data to HDFS.

```
sqoop import --connect jdbc:mysql://127.0.0.1/Project --
username root --password mypass --table olympics_History -
-target-dir /sqoop/Project/olympics_Histoy
```

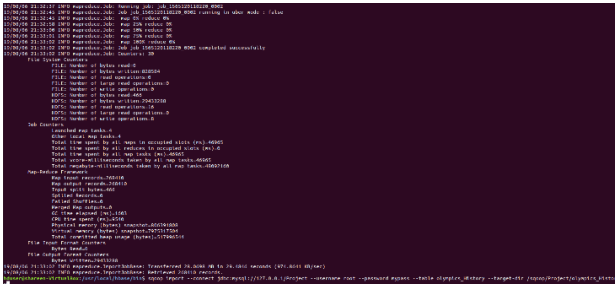


Figure 5: Data Import Command used from MySQL to HDFS

c) Hadoop and HDFS

Hadoop and HBase have to be up and running in order to process business logic since they are main data storage used for the project.

Below figure shows the service required to start before moving forward with any tasks .the ./start-dfs.sh, start-yarn.sh and ./start-hbase.sh are scripts executed to getting started with the following applications.

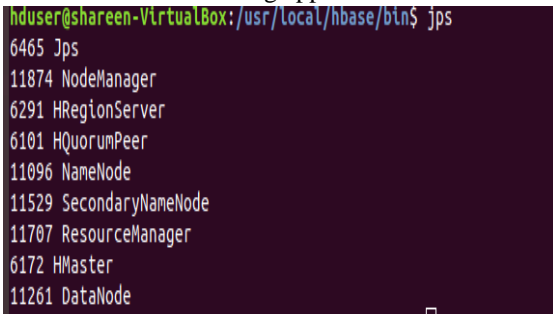


Figure 6: the service required to up and run

d) Pig MapReduce

The Data from HDFS is then loaded into Pig and after the process the business logic .the processed output is uploaded directly into HBASE with the help of pig script. The Figure below shows a simple pig script to process business logic.

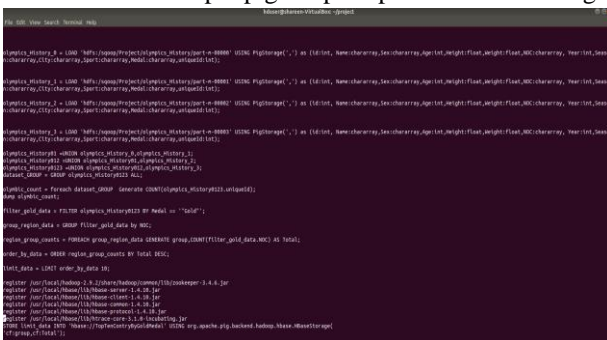


Figure 7: Sample pig script to load from HDFS and upload to HBase after processing

The above script will be executed with the help of parent automation shell script. The screenshot is attached below

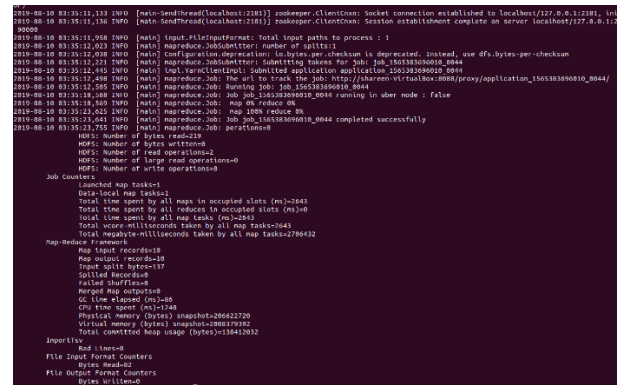


Figure 8: the output of sample pig script

After execution of pig MapReduce, the HBase will be loaded with resultant data as shown below.

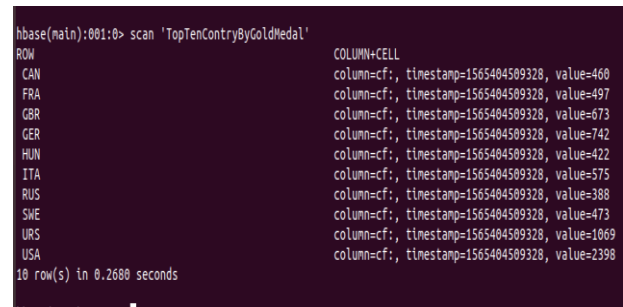


Figure 9: output loaded in the HBase table after the execution of pig script

e) Hive MapReduce

The hive also processes similar to the pig Script. The data from HDFS is loaded and the business logic is processed the result will be uploaded to HBase as before. the screenshot below shows a sample hive script and its output triggered from the automation script.



Figure 10: sample hive script

Figure 11: the output of sample hive script
After completion hive MapReduce task the output will be loaded into HBase as shown below.

The java code is written for reading the Input directory and processing the business logic and uploading output into HBase. The java is tested and exported as a jar and then integrated it with the shell script. with the help of following the command, the task is executed, and the result is uploaded into HBase. the screenshots of sample code, MapReduce output, and HBase table scan are attached below.

Figure 14: the output of the java MapReduce from shell script integrated with java jar

The HBase is NoSQL DB used to store all MapReduce data from pig hive and java MapReduce. The data from HBase is finally extracted with the help java into CSV to visualize business logic with the help of tableau or Power BI. All table used to process the business logic is shown below.

Figure 12: output loaded in the HBase table after the execution of pig script

Figure 13: Sample Code for mapper and reducer in java

Figure 15: Sample Code for insertion of Data using Java


```

hbase(main):003:0> list
TABLE
AVGheightVsWeight_olympic
GenderWiseOlympicMedalYearly
MedalWiseSport_olympic
SeasonByCity_olympic
SeasonalYearly_olympic
SentimentVsMedalinUSA
Top10Goldholders_olympic
TopTenContryByGoldMedal
customers
9 row(s) in 0.0510 seconds

```

Figure 16: The list of HBase Table used to store the solution of Business queries

D. Technologies Used and Justification for choosing them

The following section briefs the technologies used for the project and the reason why we chose them.

Apache Hadoop:

According to S. Loebman et al. (2009), The Hadoop is an Open source implemented in java in the form of MapReduce without any predefined schema used for data Storage. The

Hadoop MapReduce:

It is a Framework implemented in java for distributed Computing having two modules. It uses divide and conquers method to handle big data in parallel which increases the performance drastically. the mapper will collect the input splits and pass to the reducer. A. Bhardwaj et al. (2015).

Apache Pig

It is a high-level programming language provide parallel

Sqoop

The Sqoop is a tool which used to transfer data from Hadoop distributed system to Relational Databases and vice versa. It considered being the fast and reliable tool for data transfer between Hadoop and MySQL here P. M. Bante et al. (2017). hence Sqoop is considered

MySQL:

The structured data are commonly processed and stored using RDBMS such as MySQL. They are user-friendly reliable, efficient and flexible hence considered for the data input to

R:

We used R for preprocessing the datasets in Unix environment dues to its flexibility.

```

Task3Mapper.java Task3Reducer.java InsertData.java RetrieveData_AVGheightVsweigh
1 package extract.csv.hbase;
2 import java.io.FileWriter;
10
11 public class RetrieveData_AVGheightVsweight{
12
13     public static void main(String[] args) throws IOException, Exception{
14
15         // Instantiating Configuration class
16         Configuration config = HBaseConfiguration.create();
17
18         // Instantiating HTable class
19         HTable table = new HTable(config, "AVGheightVsWeight_olympic");
20         FileWriter wc=new FileWriter("AVGheightVsWeight_olympic.csv");
21         wc.write("avg_height,avg_weight,sport\n");
22         for(int i=1;i<68;i++)
23         {
24             // Instantiating Get class
25             Get g = new Get(Bytes.toBytes(i+""));
26
27             // Reading the data
28             Result result = table.get(g);
29
30             // Reading values from Result class object
31             byte [] value = result.getValue(Bytes.toBytes("cf"),Bytes.toBytes("height"));
32
33             byte [] value1 = result.getValue(Bytes.toBytes("cf"),Bytes.toBytes("weight"));
34
35             byte [] value2 = result.getValue(Bytes.toBytes("cf"),Bytes.toBytes("sport"));
36
37             // Printing the values
38             String height = Bytes.toString(value);
39             String weight = Bytes.toString(value1);
40
41             String sport = Bytes.toString(value2).replace("\\", "");
42             String out = height+","+weight+","+sport;
43             wc.write(out+"\n");
44
45

```

Figure 17: Sample Code for d extraction of data in the form of CSV in java from HBase

Distributed computing is supported by the Hadoop with the help of HDFS. This MapReduce programming model First introduced in GFS in 2003 and later an Opensource version of GFS developed by Doug Cutting, a researcher in yahoo. This programming model archive high performance due to its parallel processing A. Bhardwaj et al. (2015).The data on Hadoop is replicated and is built mainly for fault tolerance. This feature of Hadoop with efficient processing of real-time data without failure considered for this research.

processing by providing MapReduce. Pig Uses a procedural language known as pig Latin intern reduces the complexity.it is also developed at yahoo A. Bhardwaj et al. (2015).

Apache Hive

According to T. A. Ashwitha et al (2017), Initially, the hive was used by the Facebook users build on top of Hadoop. The hive will convert the query to MapReduce and these tasks will be executed with the help of an interpreter in parallel. Hence the research considered the Hive as well the NoSQL Databases

Apache HBase

It is an Open source NoSQL database built on top of HDFS and Hadoop. We will be storing all our MapReduce outputs into HBase tables.

Java:

The java is a high-level object-oriented programming language. Which will be capable of processing more complex queries easily. hence java is considered for the research.

Power BI

Developed by Microsoft provide excellent visualization

Tableau:

This Tool also used for the visualization for our research.

Features	Apache Pig	Apache Hive	Apache Sqoop	Apache HBase	Apache Zookeeper	Apache Flume
Developed By	Yahoo	Facebook	Cloudera	Apache Software Foundation	Yahoo	Cloudera
Available	Open-Source	Open-Source	Open-Source	Open-Source	Open-Source	Open-Source
Language supported	PigLatin	SQL-like language called HiveQL, or HQL.	MYSQL, Microsoft SQL server, PostgreSQL, IBM DB2.	Java	Java and C	Java
When to Use	For data processing on Hadoop clusters.	For analytical purposes.	When there is need to import and export data from RDBMS to Hadoop	When we need random, read/write access to our Big Data	For Distributed Applications.	For moving large amount of data to a centralized data store.
Data Structure it operates on	Complex, nested	Apache Derby database	Simple	NOSQL	Kafka data structures	Simple
Schema	Optional	Required	Optional	Required	Required	Required
External file support	Yes	Yes	Yes	No	Yes	Yes
Required Software	Java 1.6 or above is supported.	Hive version 1.2 above require Java 1.7, Hadoop version 2.x preferred.	No such requirements.	JDK version 1.7 Recommended.	JDK 6 or greater, 2 GB of RAM, Three ZooKeeper servers is the minimum recommended size	Java 1.7 Recommended, sufficient memory and disk space for source, channel, and sink.
Event Driven	No	No	No	No	No	Yes
Companies using	Yahoo	Facebook, Netflix	Yahoo, Amazon	EBay, Yahoo, TrendMicro, and Facebook etc.	Rackspace, Yahoo etc.	Yahoo, Google etc.
Used for	For processing of large data set present in Hadoop cluster	Use for effective data aggregation method, adhoc querying and analysis of huge volumes of data.	To transfer data between Hadoop and Relational databases.	To provide quick random access to huge amount of structured data.	To Provide centralized control for synchronization across the Hadoop cluster.	For moving streaming web log data into HBase.

Table 1: The Comparison of emerging Technologies in Bigdata

IV. RESULTS

This section discussed the solution obtained for the business queries mentioned in section I-A

- 1) Which are the top 10 countries with most Olympic gold medals?

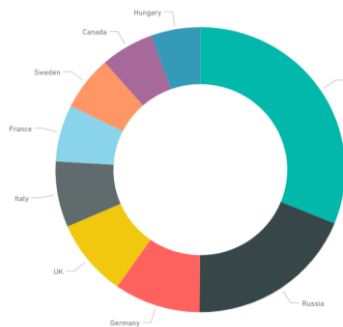


Fig. 1. Top Ten countries with Olympic gold medal

Fig.1 shows Top Ten country with Olympic gold medal. It is found that the US is the country having most of Olympic gold medal. while Russia, Germany, UK, and Italy are in Top 5.

- 2) Comparison of women Olympians who won the gold medal with men Yearly?

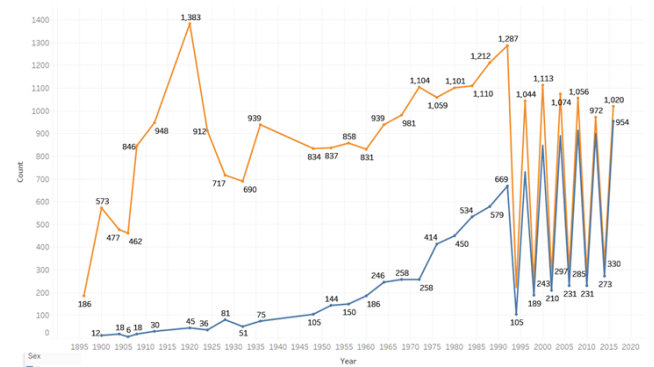


Fig. 2. Gold medal-winning gender-wise Yearly

Fig.2 shows that since 1895 onwards men dominate the Olympics till 21st century. In the 2016 Olympics, we can see that men having 1020 gold medal while women received 954 gold medals. That is an impressive change of women empowerment in the Olympics.

3) How are sentiments affecting the US Performance in Olympics?

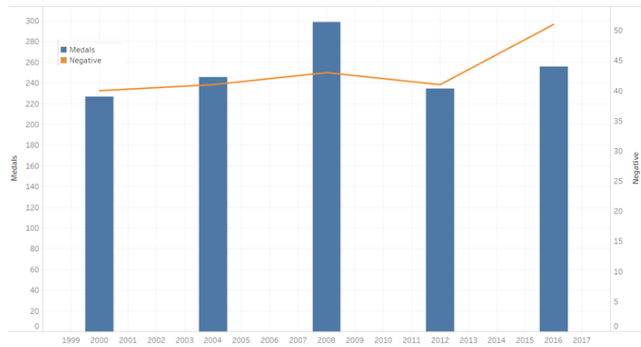


Fig. 3. Americans sentiments vs medals performance

Fig.3 shows that Americans interest towards the Olympics is decreasing (i.e. is negative sentiments increasing) as 2016 ever time high while the performance of the US in Olympics is also decreasing.

4) Which game in the Olympics has the highest participation and which one has the lowest?

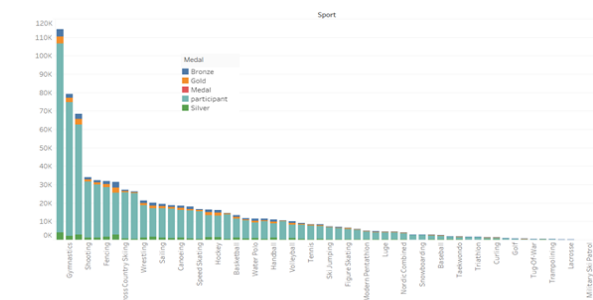


Fig. 4. Game-wise participation

Fig.3 shows that athletics gymnastics, swimming and shooting are the most popular games in the Olympics while Basque pelota, Roque, Jeu De Paume, Racquets and Motorboating are fewer known games in the Olympics.

5) Is there any difference in season-wise participation in Olympics?

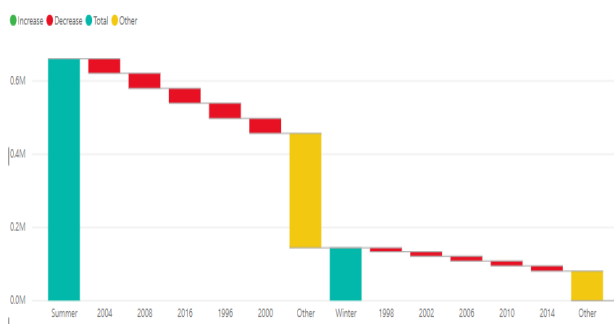


Fig. 5. Seasonal participation

Fig 5 shows that the Summer Olympics seems to have more participation when compared to winter

6) Which game demands the tall and which one demands thin people in the Olympics?

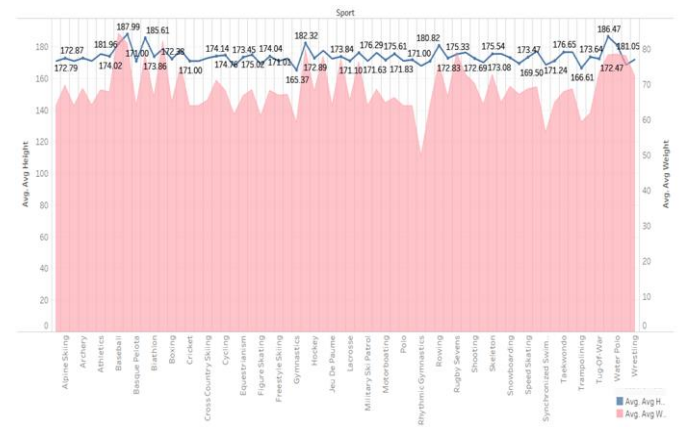


Fig. 6. Height and weight game-wise.

Fig 6 shows that the Basketball, Volleyball, Beach Volleyball, Handball, Baseball, etc. demands tall people while Thin people play well in Rhythmic Gymnastics, Synchronized Swimming, Gymnastics, Trampolining and Figure Skating, etc.

7) Top 10 gold medal hunters in Olympics history?



Fig. 7. Top Ten gold medal hunters in Olympics

Fig7 shows that Raymond Clarence Ray Ewry is having a total of 30 gold medals in hand. Paavo Johannes Nurmi, Larysa Semenivna Latynina, Frederick Carlton Carl Lewis, Mark Andrew Spitz has 27 each. While Usain St. Leo Bolt is only having 24 medals

8) Most preferred cities hosting the Olympic game season-wise?



Fig. 8s. Heat map for season wise popular hosting city.

Fig8 shows that Landon is the most preferred summer Olympics host while Sochi is the most preferred winter

V. CHALLENGES AND LIMITATIONS

The development of the system involves several challenges. The image to text extracting of unstructured data was challenging due to the image quality and size. The image is rescaled in order to achieve the desired output. many issues faced during the project development such as hive, and pig version was not compatible with HBase to upload MapReduce data .hence multiple times software has to reinstall in order to compatible with HBase along with upgrading of databases. The Extraction of CSV from HBase was also challenging the inbuilt tool gives garbage along with data hence a java program has to implement to extract. The most challenge faced is integrating everything into a single script. The other issue faced is MySQL loading a few rows were skipped due to invalid data on those cells.

REFERENCES

- [1]G. Ming and X. Man, "Men Athletics Results of Value Analysis," 2011 International Conference on Future Computer Science and Education, Xi'an, 2011, pp. 369-374.
- [2]Binhui Wang, Chuxiang Chen and Ke Liu, "The research of Olympic medal ranking methods," 2011 2nd International Conference on Artificial Intelligence, Management Science and Electronic Commerce (AIMSEC), Dengleng, 2011, pp. 7406-7410.
- [3] C. Thirumalai, S. Monica and A. Vijayalakshmi, "Heuristics prediction of Olympic medals using machine learning," 2017 International conference of Electronics, Communication, and Aerospace Technology (ICECA), Coimbatore, 2017, pp. 594-597.
- [4] T. A. Ashwitha, A. P. Rodrigues and N. N. Chiplunkar, "Movie Dataset Analysis Using Hadoop-Hive," 2017 2nd International Conference on Computational Systems and Information Technology for Sustainable Solution (CSITSS), Bangalore, 2017, pp. 1-5.
- [5]A. Bhardwaj, Vanraj, A. Kumar, Y. Narayan and P. Kumar, "Big data emerging technologies: A CaseStudy with analyzing twitter data using apache hive," 2015 2nd International Conference on Recent Advances in Engineering & Computational Sciences (RAECS), Chandigarh, 2015, pp. 1-6.
- [6] D. Patel, X. Yuan, K. Roy, and A. Abernathy, "Analyzing network traffic data using Hive queries," SoutheastCon 2017, Charlotte, NC, 2017, pp. 1-6.

VI. CONCLUSION AND FUTURE WORK

The Hadoop MapReduce system developed for analyzing the Olympics insight is achieved its objective by finding a solution to its business queries mentioned in section 1. The Input data is processed with the help of R and unstructured data is extracted with OCR. The Sqoop is used to import data between MySQL and HDFS. different MapReduce was used to solve the business queries. They were Java, hive, and pig MapReduce. The result of MapReduce is stored in HBase. The HBase schema is extracted into CSV with the help of java to visualize. This demonstrates the Bigdata framework capabilities over a large set of data using MapReduce. The future work includes analyzing the time complexity between each frameworks hive, pig and java. The other works looking forward is analyzing similar frameworks such as a spark, Apache Flink, etc.

- [7] W. Xu et al., "Supporting large scale connected vehicle data analysis using HIVE," 2016 IEEE International Conference on Big Data (Big Data), Washington, DC, 2016, pp. 2296-2304.
- [8] S. Loebman, D. Nunley, Y. Kwon, B. Howe, M. Balazinska and J. P. Gardner, "Analyzing massive astrophysical datasets: Can Pig/Hadoop or a relational DBMS help?," 2009 IEEE International Conference on Cluster Computing and Workshops, New Orleans, LA, 2009, pp. 1-10.
- [9] P. M. Bante and K. Rajeswari, "Big data analytics using Hadoop map-reduce framework and data migration process," in 2017 International Conference on Computing, Communication, Control and Automation (ICCUBEA). IEEE, 2017, pp. 1–5.
- [10] A. Hennig et al., "Big social data analytics of changes in consumer behavior and opinion of a TV broadcaster," 2016 IEEE International Conference on Big Data (Big Data), Washington, DC, 2016, pp. 3839-38
- [11] Li, Y., Liang, L., Chen, Y., Morita, H., 2008. Models for measuring and benchmarking Olympic achievements. Omega International Journal of Management Science 36, 6, 933–940.
- [12]Li, Y., Chen, Y., Liang, L., & Xie, J. (2012). DEA models for extended two-stage network structures. Omega, 40(5), 611–618.
- [13] Wu, Jie, Liang, Liang, and Yao Chen. 2009. "DEA Game Cross-Efficiency Approach to Olympic Rankings," Omega, 37, 909-918