



BIG DATA FOR ANALYSIS OF COMMONLY CULTIVATED CROPS YIELD PRODUCTION IN ANDHRA PRADESH

P.Bhargavi¹, R.P.L.Durgabai², S.Jyothi³

Department of Computer Science,
Sri Padmavati Mahila Visvavidyalayam,
Tirupati, A.P, India

pbhargavi18@yahoo.co.in,
poonamramchandra@gmail.com,
jyothi.spmvv@gmail.com

June 22, 2018

Abstract

Now-a-days as data is growing rapidly analysis of data is playing a key role for efficient decision making. Big data refers to the ability to collect and analyze the huge amounts of data that is being generated by different departments working directly or indirectly involved in agriculture . Due to lack of resources the analysis of crop yield production is very slow which leads to take poor decisions that occurs loss to farmers. The goal is to provide better solution for finding problems and to enhance the crop yield productivity of the agriculture sector. This kind of analysis can be done using big data tools to make better decisions on various crop yield productivity in Andhra Pradesh.

Keywords:Apache Pig, Crop yield, Big Data ,Agriculture

1 Introduction

Big data analytics has made such a widespread impact in the agriculture industry that its hard to pinpoint all its effects, and harder still to predict what changes it might bring. Big Data in the context of agriculture refers to new means of collecting and analyzing data generated from the farm to the end consumer. Big data is a set of methods and technologies that require new forms of combination to uncover large hidden values from large datasets that are diverse, complex, and of a mass[1].

In India Andhra Pradesh is an agriculturally important state. Agriculture Sector is playing a very significant role in state economy. In Andhra Pradesh agriculture is mostly dependent on rainfall. crop production is dependent on seasonal conditions and monsoons. The State is in third place of production in rice and groundnuts and in Second position in cotton and sunflower. The state has been implementing various schemes to address the problems of agriculture from time to time. In spite of this, agriculture in the state has been exhibiting stagnation in growth and is seeking innovative policy and technology interventions[2]. This paper presents different varieties of commonly cultivated crops in Andhra Pradesh and their normal expansion, production, crop productivity during khariff and Rabi seasons which is helpful for farmer to make better decisions in productivity of crops.

2 Big Data

To examine large and varied data sets and to uncover hidden patterns, market trends, unknown correlations, customer preferences and other useful information which can help organizations to make more informed business decisions the emergence and growth of big data analytics is essential. Traditional data ware houses are based on relational databases to structured data sets and it typically not suitable for unstructured and semi-structured data types. Big data is updated frequently and continually , as in the case of real-time data on stock trading, the online activities of website visitors or the performance of mobile applications the traditional databases and data ware houses may not able to handle the data and processing demands

of big data.

As a result, many organizations that collect, process and analyze big data turn to NoSQL databases as well as Hadoop and its companion tools, including:

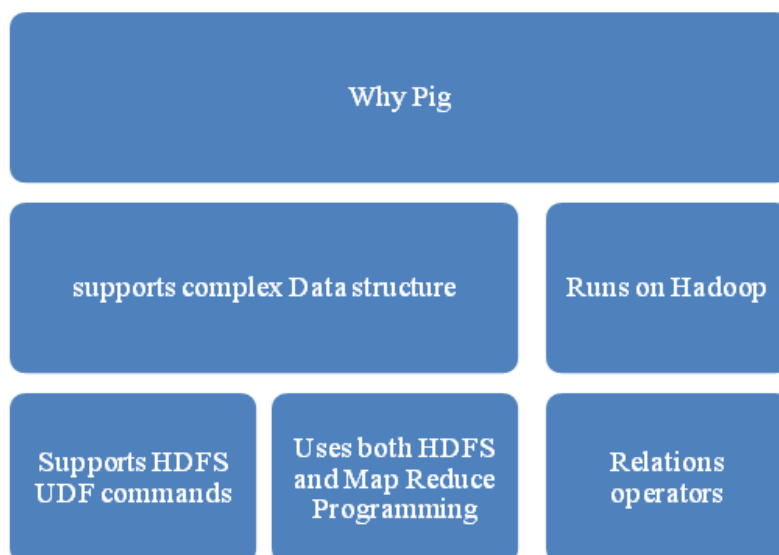
Table 1: Hadoop Tools

Tools	Description
YARN	A cluster management technology and one of the key features in second-generation Hadoop
MapReduce	A software framework that allows developers to write programs that process massive amounts of unstructured data in parallel across a distributed cluster of processors or stand-alone computers.
SPARK	an open-source parallel processing framework that enables users to run large-scale data analytics applications across clustered systems.
Hbase	a column-oriented key/value data store built to run on top of the Hadoop Distributed File System (HDFS).
Hive	an open-source data warehouse system for querying and analyzing large datasets stored in Hadoop files
Kafka	distributed publish-subscribe messaging system designed to replace traditional message brokers
Pig	an open-source technology that offers a high-level mechanism for the parallel programming of MapReduce jobs to be executed on Hadoop clusters.

2.1 Apache Pig

Apache pig is generally used with Hadoop, and all the data manipulation operations is possible in Hadoop using pig [3]. It is a tool/platform which is used to analyze larger sets of data representing them as data flows. It is an abstraction over map reduce .Apache Pig is a high-level language platform developed to execute queries on huge datasets that are stored in HDFS using

Apache Hadoop. It is similar to SQL query language but applied on a larger dataset and with additional features[4]. The language used in Pig is called Pig Latin. It is very similar to SQL. It is used to load the data, apply the required filters and dump the data in the required format. It requires a Java runtime environment to execute the programs. Pig converts all the operations into Map and Reduce tasks which can be efficiently processed on Hadoop. It basically allows us to concentrate upon the whole operation irrespective of the individual mapper and reducer functions.



Pig is having features like Rich set of Operators ,Ease of programming, Optimization opportunities, extensibility, User Defined Functions, Handling all kinds of data. Pig consists of a collection of rich set of operators in order to perform operations such as Join, filter, sort and many more. Pig Latin is Similar to SQL and hence it becomes very easy for developers to write a pig Script. The execution of the task in Apache Pig gets automatically optimized by the task itself, hence the programmers need to focus only on the semantics of the language. By using the existing operators, users can develop their own functions to read, process and write data. Pig provides the facility to create User - Defined Functions in other programming languages such as Java

and invoke or embed them in Pig Scripts. It analyzes all kinds of data, both structured as well as unstructured. It stores the results in HDFS. Apache Pig is generally used by data scientists for performing tasks involving ad hoc processing and quick prototyping [10]. Pig is used to process huge data sources such as web logs, To perform data processing for search platforms, To process time sensitive data loads. Pig Supports complex data structures and runs on Hadoop. It uses both HDFS and Map Reduce Programming and supports HDFS and UDF commands.

3 Experimental Set up and Results

In this experiment setup Edureka virtual Machine Hadoop 2.7.3 infrastructure with configuration single node has 6GB RAM and i5 CPU . Data is collected from Acharya N.G. Ranga Agriculture University and Agriculture Department of Government. In Andhra Pradesh during khariff and rabi seasons total 28 varieties of crops will be cultivated in 63.64 lac hectares land expansion. Among those crops few important crops are rice, maize, pulses, groundnut, cotton ,chilly, tobacco, sugarcane. The present Dataset consists about 28 different varieties of Crops includes food grains, millets, pulses, total food grains, total oilseeds, other crops like tobacco, chilly, cotton, onions, fibre crop, sugarcane, turmeric, and their total land expansion in state, total yield, total crop yield per hectare during khariff and Rabi seasons. The land expansion is in lakh hectares, crop productivity is in kilo gram per hectare, yield is in lac hectares. Data description includes Sno, Cropname, land expansion , yield, Crop productivity during khariff and rabi seasons. Apache Pig is used to analyze the data and queries have been generated. Pig programs can be executed in three methods. write a pig script file containing all the commands and execute it from the command line. Interactive shell grunt can be used to execute the commands line by line. The required commands can be executed by extending the PigRunner class. The method applied is interactive shell grunt to execute commands line by line.[17]

Now take a query to load the AP crop data, first open the terminal and give pig -x local to store the data in a local mode and when the grunt shell is open give the following command to

```
load the AP crop data.[5] food=
load'/home/edureka/desktop/applantdata.txt' using pig storage
('/t');
```

To find the food grains normal expansion in lac hectares
 grunt>filtergrain=filter food by \$0<7 ;
 grunt>dump filtergrain;
 dump is the used to run the command and give the output

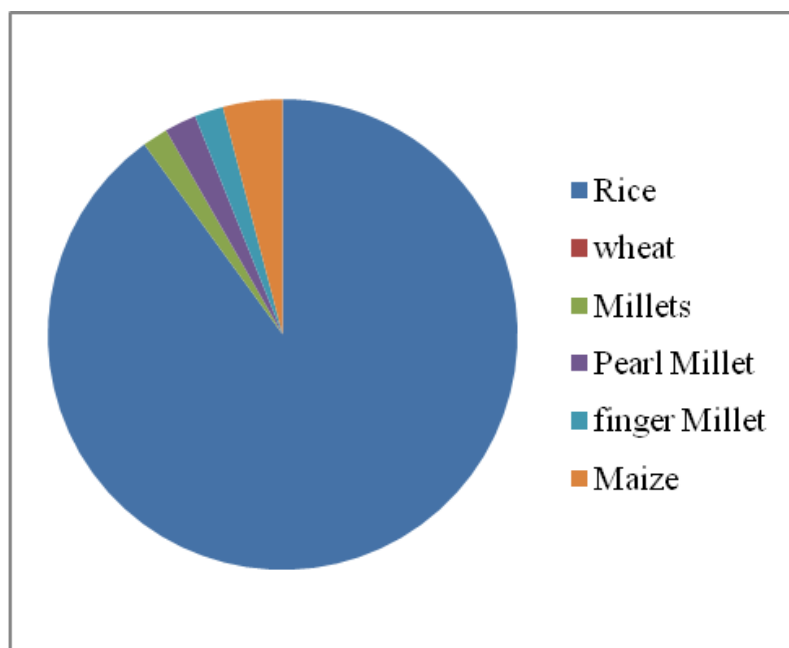


Figure 1: Food grains Expansion in lakhs hectares

The above graph gives the complete result of food grains like rice, wheat, millets, pearl millets, finger millets, maize and only millets, maize is having high expansion of land than other food grains. The normal expansion of Rice is 16.42 lakh hectares, wheat is 0 hectares, millets is 0.32 lakh hectares, pearl millet is 0.40 lakh hectares, finger millet is 0.36 lakh hectares and maize is 0.75 lakh hectares. with this analysis it is easy for a farmer to estimate how much land expansion is required for each crop.

To show all the cereals area of expansion
 grunt>filtercereal = filter food by \$0>7 and \$0<14;
 grunt>dump filtercereal;

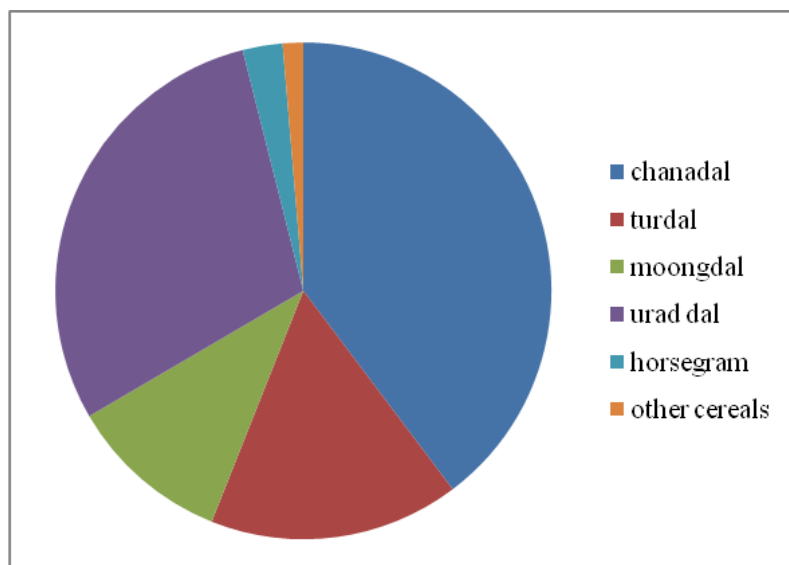


Figure 2: Cereals area of expansion in lakh hectares

The above graph gives the output of cereals and among all the cereals chanadal area of expansion is high. The normal expansion of all cereals is chanadal- 5.04 lakh hectares , turdal- 2.07 lakh hectares, moongdal-1.34 lakh hectares ,uraddal - 3.74 lakh hectares,horsegram - 0.33 lakh hectare,other cereals - 0.17 lakh hectares.

Find all the oil seeds
 grunt>oilseeds= filter food by \$0>20 and \$0<23;
 grunt>dump oilseeds;

The above query gives the output of Oil seeds and palmoil/other oil seeds expansion is more than castroil. This query shows the Normal expansion of oil seeds like castor oil and

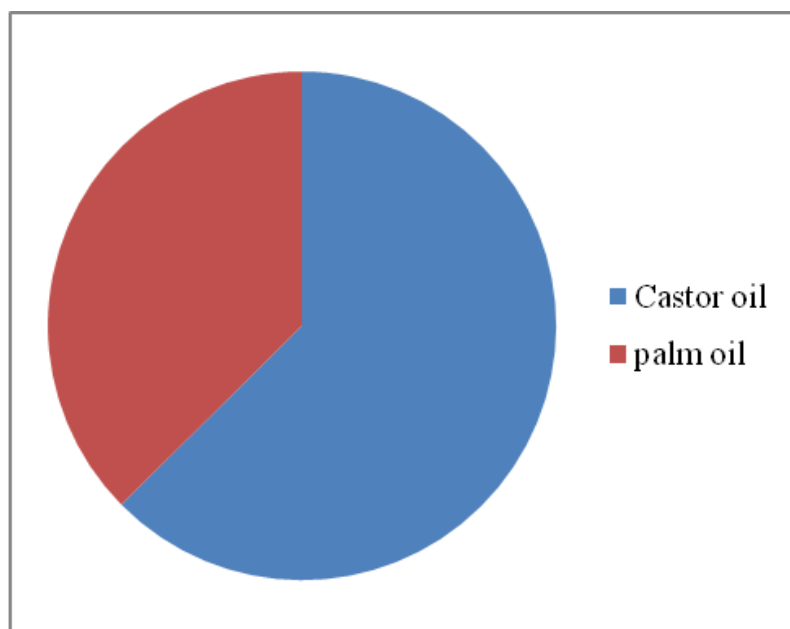


Figure 3: Expansion of Oil seeds

palmoil/other oil seeds. The normal expansion of castor oil is 0.87 lakh hectares, in previous years it is 0.49 lakh hectares, 0.47 lac hectares, the crops productivity in kilo gram per hectare in previous years is 527 lakh tonnes, 574 lakh tonnes is 0.26 lakh hectares, 0.27 lakh hectares. The palm oil and other oil seeds normal expansion is 0.52 lakh hectares, and expansion in previous years is 0.55 lakh hectares, 0.61 lakh hectares, crop productivity is 22319 kilo gram per hectare, 21419 kilo gram per hectare, The yield is 12.34 lakhs tonnes and 13.07 lakhs tonnes, in both khariff and Rabi seasons.

show production of rice from 2013 to 2015

```
grunt>riceproduction= foreach food generate $1 as meal,$8 as
production;
grunt>ricefilter=filter riceproduction by meal=='Rice';
grunt>dump ricefilter;
```

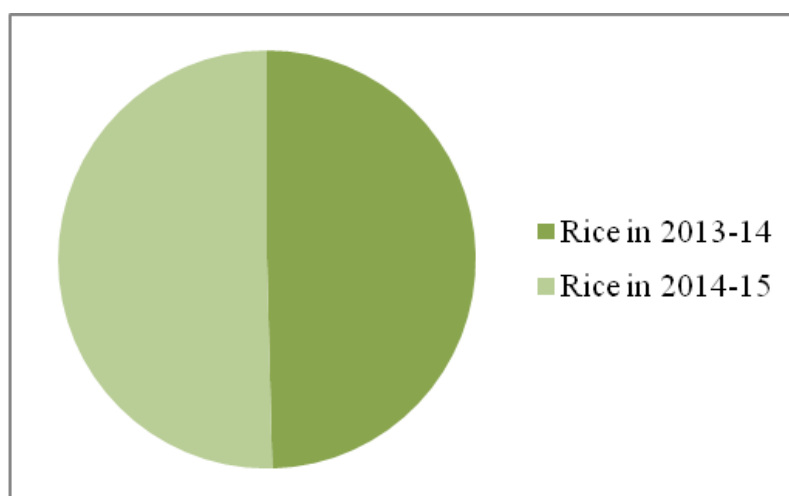


Figure 4: Production of Rice from 2013 to 2015

The above query display the production of rice during the year 2013-15. During 2013-14 the production of rice is 79.94 lakh tonnes and in 2014-15 the production of rice is 81.4 lakh tonnes.

Show normal expansion of maize.

```
grunt>maize= foreach food generate $ as mean,$2 as Normal;
grunt>maize expansion=filter maize by meal=='Maize';
grunt>dump maize expansion;
```

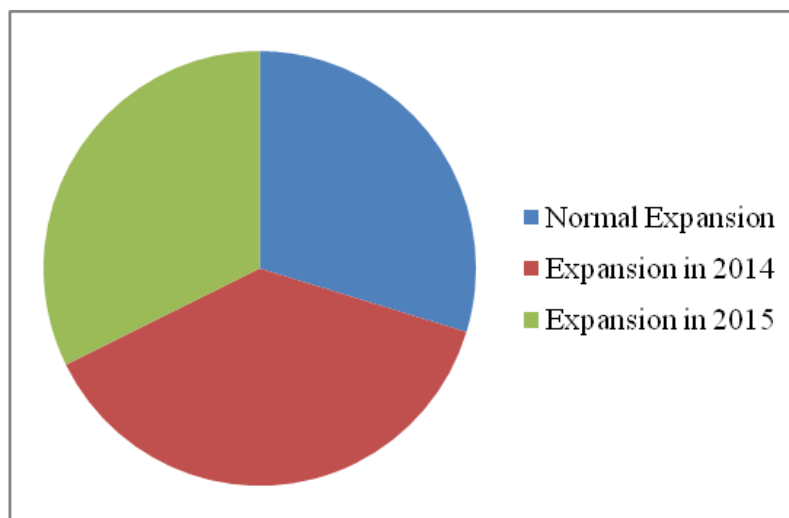


Figure 5: Normal Expansion of Maize

The above query shows the normal area expansion of Maize which is 2.76 lakh hectares and the normal expansion of Maize in 2013-14 is 3.52 lac hectares and in 2014-15 is 3.00 lac hectares. The crop during 2013-14 is 6286 kilo gram hectares, 2014-15 is 6260 kilo gram hectares. The Production of Maize in 2013-14 is 22.14 lakh tonnes and 2014-15 is 18.78 lakh tonnes.

show total coarse grains.

```
grunt> coarsegrain=group filtergrain all;
grunt>dump coarse grain;
grunt>grains_count=foreach          coarsegrain          Generate
COUNT(filtergrain);
```

The above query gives the total coarse grains production during the seasons of khariff and Rabi. There are total six coarse

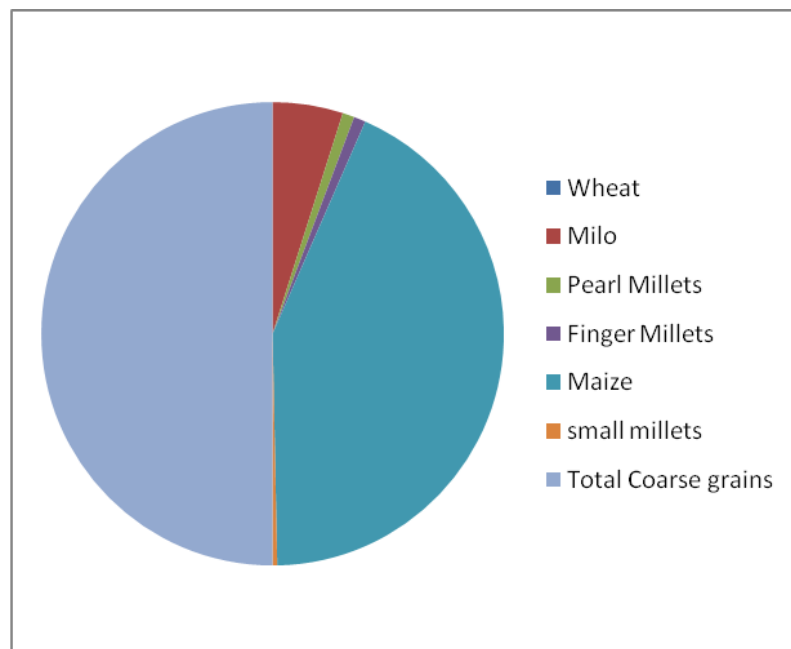


Figure 6: Coarse Grains Production

grains they are wheat,milo,pearl millet,finger millet,maize,small millets. The production of Rice is high among all the other grains. wheat 0,Milo 2.14,pearl millet 0.36 , finger millet 0.36, maize 18.78,small millets 0.14and total production of coarse grains is 21.78 lakh tonnes.

```
show total food grains
grunt>allfoodgrains=filter food by $0<14;
grunt>grains=dump allfoodgrains all;
grunt>count_foodgrains=foreach          grains          generate
count(allfoodgrains);
grunt>dump count_foodgrains;
```

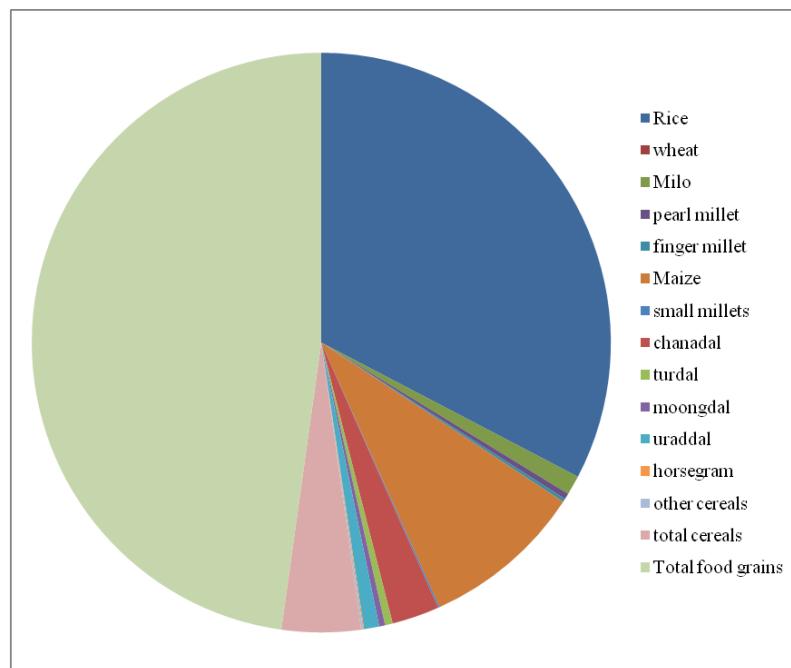


Figure 7: Total food grains production during 2013-14

The above query gives the output of total food grains and its production during khariff and Rabi seasons in 2013-14. Production of each food grain is Rice - 79.94 lakh tonnes, wheat

-0 tonnes, milo- 2.65 lakh tonnes, pearl millets - 0.82 lakh tonnes, finger millets - 0.43 lakh tonnes, maize - 22.14 lakh tonnes, small millets - 0.21 lakh tonnes, chanadal -6.48 lakh tonnes, turdal -1.04 lakh tonnes, moongdal- 0.82 lakh tonnes, uraddal- 2.07 lakh tonnes, horsegram-0.15 lakh tonnes, othercereals-0.24 lakh tonnes, total cereals-10.79 lakh tonnes, total food grains production-116.98 lakh tonnes

Groundnut expansion

```
grunt>Groundnut= foreach food generate $1 as nut,$6 as exp,$7
as expansion;
grunt>ground=filter groundnut by nut=='Groundnut';
grunt>dump ground;
```

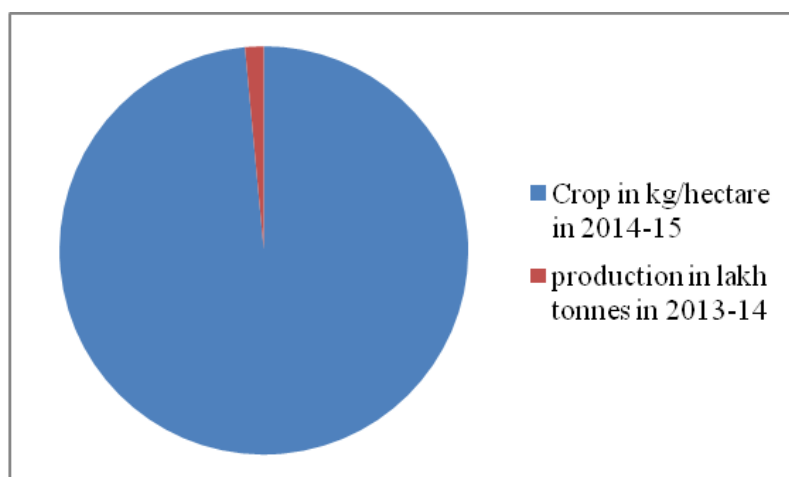


Figure 8: Ground nut expansion

The above query shows the ground nut crop kilo gram per hectare in 2014-15, the crop yield in 2014-15 is 585 kg/hectare and the production of groundnut in 2013-14 is 8.81 lakh tonnes and 5.11 lakh tonnes in 2014-15.

4 Conclusion

Agriculture data has entered the world of big data as agriculture has been transformed into a technological world. Big Data analytics is a technique that help us in a decision making. To analyze the varieties of crops production like food grains, oilseeds, cereals, millets, coarse grains, maize, groundnut in AP region, big data analytics framework has been developed that provided solutions by using pig . Based on evidence from original data the proposed work is helpful for researchers and officers for recommending solutions, the work is helpful to make decisions as on which season a particular crop in AP region can produce high and the area of expansion of each crop and to analyze the total crop production based on previous years.

References

- [1] T.GiriBabu,Dr.G.Anjan Babu,March 2016."Big Data Analytics to Produce Big results in the Agriculture Sector", International Journal of Advanced Research in Biology Engineering Science and Technology,vol2,ISSN 2395-695x.
- [2] C.A.RamaRao, K.Karimulla, Sreenath Dixit,Y.S.RamaKrishna,K.Ravishankar,2008."Performance of Agriculture in Andhra Pradesh - A Spatial and Temporal Analysis", Central Research Institute for Dry land Agriculture, ISBN : 978 - 81-904360-3-8.
- [3] Xu Y,Zhou W,Cui B,Lu L. July 2015. "Research on performance optimization and visualization tool of hadoop". proceedings of 10th International conference on Computer Science and Education;
- [4] Ammar Fua,Alva Erwin,Heru purnomo ipung,September 2014."Processing performance on Apache pig,Apache hive, MySQL cluster",IEEE Publisher,DOI.10.1109/ICTS.2014.7010600.
- [5] Raghu garg,Himanshu Aggarwal, August 2016." Big Data Analytics recommendations solutions for crop disease using

- hive and hadoop platform", Indian Journal of Science and Technology, vol9(32), ISSN:0974-5645.
- [6] Bankim L.Radadiya, Parag Shukla,2017. "Analyzing Varieties of Agriculture Data using Big Data Tools Pig", Oriental Journal of Computer Science and Technology, vol10, ISSN:0974-6471.
 - [7] Sarika Rathi, January 2017. "A brief study of Big data Analytics using Apache Pig and Hadoop Distributed File System", International Journal of Advanced Research in Computer Engineering & Technology, vol(6), Issue 1, ISSN:2278-1323
 - [8] Sanjeev Dhawan, Sanjay Rathee, May 2013. "Big Data Analytics using Hadoop Components like Pig and Hive," American International Journal of Research in Science, Technology, Engineering and Mathematics. pp.88-93
 - [9] S. Vinila Kumari, Dr.P.Bhargavi, U.Subhashini, March 2016. "Role of Big Data Analytics in Agriculture", IJCSME-SCSMB, ISSN-2349.
 - [10] Kranti bansal, Priyanka Chawla. "A study of Big Data Analysis using Apache Pig", International Journal for IJCTA pp8665-8672.
 - [11] Swarna C, Zahid Ansari, August 2017. "Apache Pig- A Data flow FrameWork based on Hadoop Map Reduce, International Journal of Engineering Trends and Technologies, vol(50), ISSN:2231-5381.
 - [12] Seema Acharya, Subhashini Chellapan, 2015. "Big Data and Analytics" Wiley Publication
 - [13] Dr. Birendra Goswami, Pradip Kumar Chandra "The Evolution Of Big Data As A Research And Development" International Journal of Scientific Research and Engineering Studies (IJSRES) Volume 2 Issue 3, March 2015 ISSN: 2349-8862

- [14] Agarwal,Shafali,Zeba Khanam,2015. "Map Reduce: A Survey Paper on Recent Expansion", International Journal of Advanced Computer Science and Applications,6.8,209-215.
- [15] Bhosale,Harshawardhan S,Devender P.Gadekar,2014."A Review Paper on Big Data and Hadoop". International Journal of Scientific and Research Publications 4.10.
- [16] Chavan,Ms Vibhavari, Rajesh N.Phursule,2014."Survey paper on Big Data",International Journal of Computer Science Information Technology ,5.6,ISSN:7932-7939.
- [17] Samak. Taghrid,Daniel Gunter,Valerie Hendrix,2012. "Scalable analysis of network measurements with Hadoop and Pig." Network Operations and Management Symposium,IEEE.
- [18] Yethiraj N G,Noor Ayesha,April 2017. " A study to improve crop Yield in Agriculture using IOT and Bigdata",Adarsh Journal of Information Technology,vol(6),ISSN:2320-0340.
- [19] Laney D , Feb 2001."3D data management: Controlling data volume, velocity and variety". Meta Group Inc Application Delivery Strategies;
- [20] Chen X-W, Lin X,May 2014." Big data deep learning challenges and perspective". IEEE Access. 2:51422.
- [21] Marx V, Jan 2013. "Biology: The big challenges of big data Nature". 498(7453):25560.
- [22] Big data in agriculture. Available from: <http://www.citethisforme.com/topic-ideas/technology/Big%20Data-6678234>
- [23] Zhang H, Wei X, Zou T, Li Z, Yang G, Sep 2014." Agriculture big data: Research status, challenges and countermeasures. Proceedings of Computer and Computing Technologies in Agriculture"; China.
- [24] Schumacher A, Pireddu L, Niemenmaa M, Kallio A, Korpelainen E, Zanetti G, Heljanko K,jan 2014." Simple and

scalable scripting for large sequencing data sets in hadoop. Bioinformatics".Vol30(1).

- [25] Nordberg H, Bhatia K, Wang K, Wang Z, Sep 2013." Biopic: A hadoop-based analytic toolkit for large-scale sequence data", Oxford Bioinformatics. Vol29(33):30149.

