

All India Council for Technical Education, Government of India

| Idea / Approach Details | | | |
|--------------------------------|--|----------------------------|--------------|
| Problem Statement: | Software – Report Generation for MGNREGA | Problem ID: | MK202 |
| Technology Bucket: | Finance | Category: | Software |
| Ministry: | Rural Development | AICTE Institute ID: | 1-1366332831 |
| Team Leader: | M. Vigneshwaran | Team Name: | TEAM TYRO |

Members:

Prawinrajan N

Yamini VP

Mohamed jakkariya R

Prithiviraj K

Sarmila S

Mentors:

Mrs. R Renuga devi

dept.of IT, BIT campus

renuinfotech23@gmail.com

Mr. R Deepan Chakkaravarthy

Stack Route, Bangalore

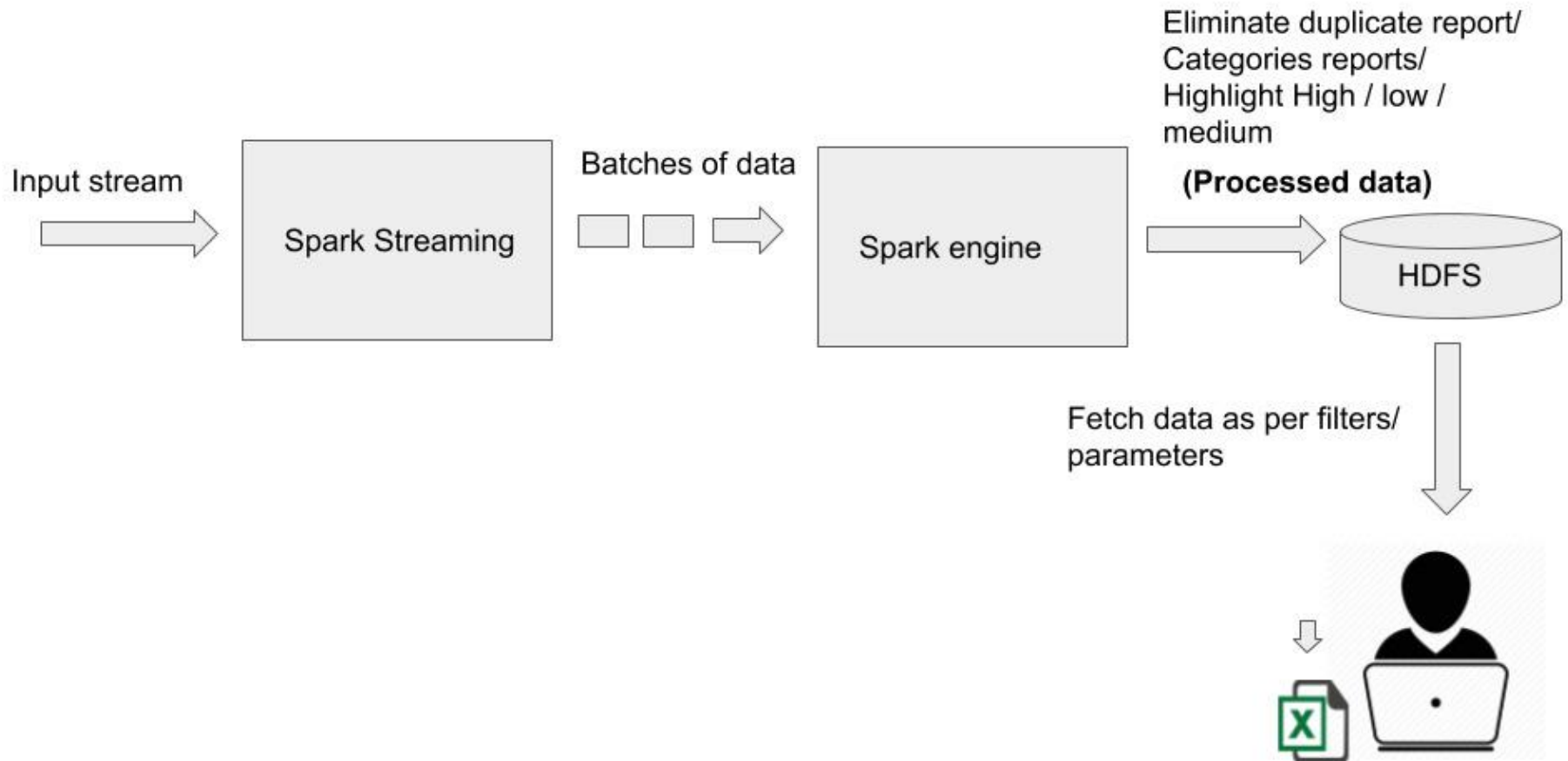
dchakravarthy09@gmail.com

objectives

Python Data Pre-processing using Spark Dataframe

- 1. Loading Data (Loading CSV file into HDFS, HDFS to Spark)
- 2. Exploring Data
 - 2.1 Understanding Dataframe Schema
 - 2.2 Obtaining summary statistics
 - 2.3 Group By and Aggregation
 - 2.4 Visualizing Data
- 3. Cleaning Data
 - 3.1 Filtering Data
- 4. Streamlining the reporting data
- 5. Eliminating the duplicate report / categorize report
- 6. Highlighting the High/Medium/Low importance reports
- 7. Generation of reports by selection required filters/parameters.

Architecture of SPARK



Screenshots of the output

Registration form



Mahatma Gandhi National Rural Employment Guarantee Act 2005
Ministry of Rural Development
Government of India



Register

Email

jack@gmail.com

Password

....

Login

No account ? [Register](#)





Mahatma Gandhi National Rural Employment Guarantee Act 2005
Ministry of Rural Development
Government of India



FINANCIAL YEAR

2020 - 2021

STATE

TAMILNADU

REPORT

R1.STAFF Registration

1. [Staff Registration Report](#)
2. [Designation wise Staff Registration Report](#)
3. [Total Number of Staff with Aadhaar and Mobile Number](#)
4. [Staff Registration and Verification Report](#)



Mahatma Gandhi National Rural Employment Guarantee Act 2005
Ministry of Rural Development
Government of India



FINANCIAL YEAR

2018 - 2019

STATE

KERALA

REPORT

R2,DIGITAL PAYMENTS

1. [Training Status](#)
2. [Location Covered for Digital Payment Training](#)



Mahatma Gandhi National Rural Employment Guarantee Act 2005
Ministry of Rural Development
Government of India



FINANCIAL YEAR

2019 - 2020

STATE

MAHARASHTRA

REPORT


R3.PROJECT

1. [Project-UNNATI - Statewise Status](#)


Generating Reports

Browser tabs: (2) WhatsApp, Meet - amr-unwd-uoi, PHYSICAL PROGRESS REPORT, mnregaweb4.nic.in/netnrega/MIS

File | C:/xampp/htdocs/sih/report_index.html



The Mahatma Gandhi National Rural Employment Guarantee Act 2005
Ministry of Rural Development
Government of India



MIS REPORTS

Kindly select Which report you want to manage!!

PHYSICAL PROGRESS

1. Progress Report
2. SC ST Employment Status
3. Employment Provided to Disabled Persons
4. Employment Provided Period wise
5. GP With No Employment

[Click here](#)

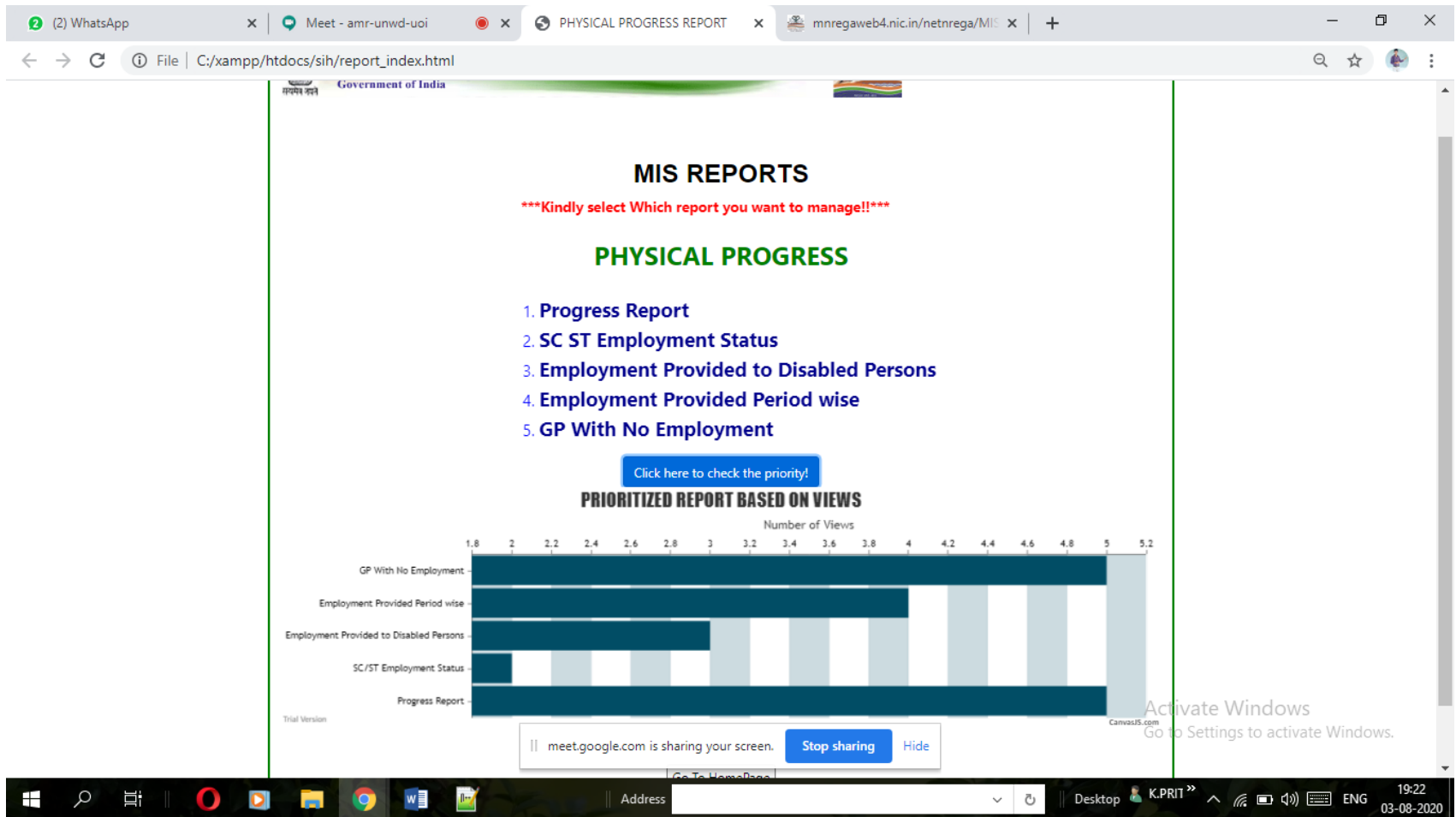
file:///C:/xampp/htdocs/sih/progressmain.html

meet.google.com is sharing your screen. [Stop sharing](#) [Hide](#)

Activate Windows
Go to Settings to activate Windows.

Desktop K.PRIT 19:20 03-08-2020

Priority reports based on times of view



Inserting function

Browser tabs: (2) WhatsApp, Meet - amr-unwd-uo, EMPLOYMENT GENERATED DUR, mnregaweb4.nic.in/netnrega/M...

File | C:/xampp/htdocs/sih/scstempinsert.html

SC ST Employment Status

District

HH issued jobcards

| | |
|--------|-------------------------------------|
| SC | <input type="text" value="2333"/> |
| ST | <input type="text" value="23444"/> |
| Others | <input type="text" value="234234"/> |
| Women | <input type="text" value="232342"/> |

HH issued Employment

| | |
|--------|-------------------------------------|
| SC | <input type="text" value="234234"/> |
| ST | <input type="text" value="23234"/> |
| Others | <input type="text" value="234234"/> |
| Total | <input type="text" value="23234"/> |
| Women | <input type="text" value="23123"/> |

No. of Persondays generated

| | |
|--------|-------------------------------------|
| SC | <input type="text" value="12312"/> |
| ST | <input type="text" value="123123"/> |
| Others | <input type="text" value="123123"/> |
| Total | <input type="text" value="1231"/> |
| Women | <input type="text" value="123123"/> |

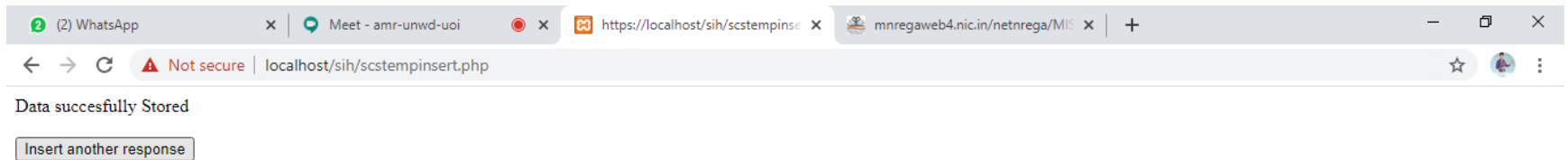
Families completed

| | |
|--------|--------------------------------------|
| SC | <input type="text" value="123123"/> |
| ST | <input type="text" value="1231231"/> |
| Others | <input type="text" value="23"/> |
| Total | <input type="text" value="23"/> |

meet.google.com is sharing your screen.

Activate Windows
Go to Settings to activate Windows.

Acknowledgement



https://localhost/sih/scstempinsert.html


meet.google.com is sharing your screen. [Stop sharing](#) [Hide](#)


Activate Windows
Go to Settings to activate Windows.

Retrieving function

(2) WhatsApp x Meet - amr-unwd-uoi x SC/ST jobcards x mnregaweb4.nic.in/netnrega/M... x

File | C:/xampp/htdocs/sih/scstempret.html

 **The Mahatma Gandhi National Rural Employment Guarantee Act 2005**
Ministry of Rural Development
Government of India



SC/ST Retrive records

***Kindly select the District and academic year**

District

Academic Year

meet.google.com is sharing your screen.

[Stop sharing](#)

[Hide](#)

Activate Windows

Go to Settings to activate Windows.

ST/ST JobCard Report 2019-2020

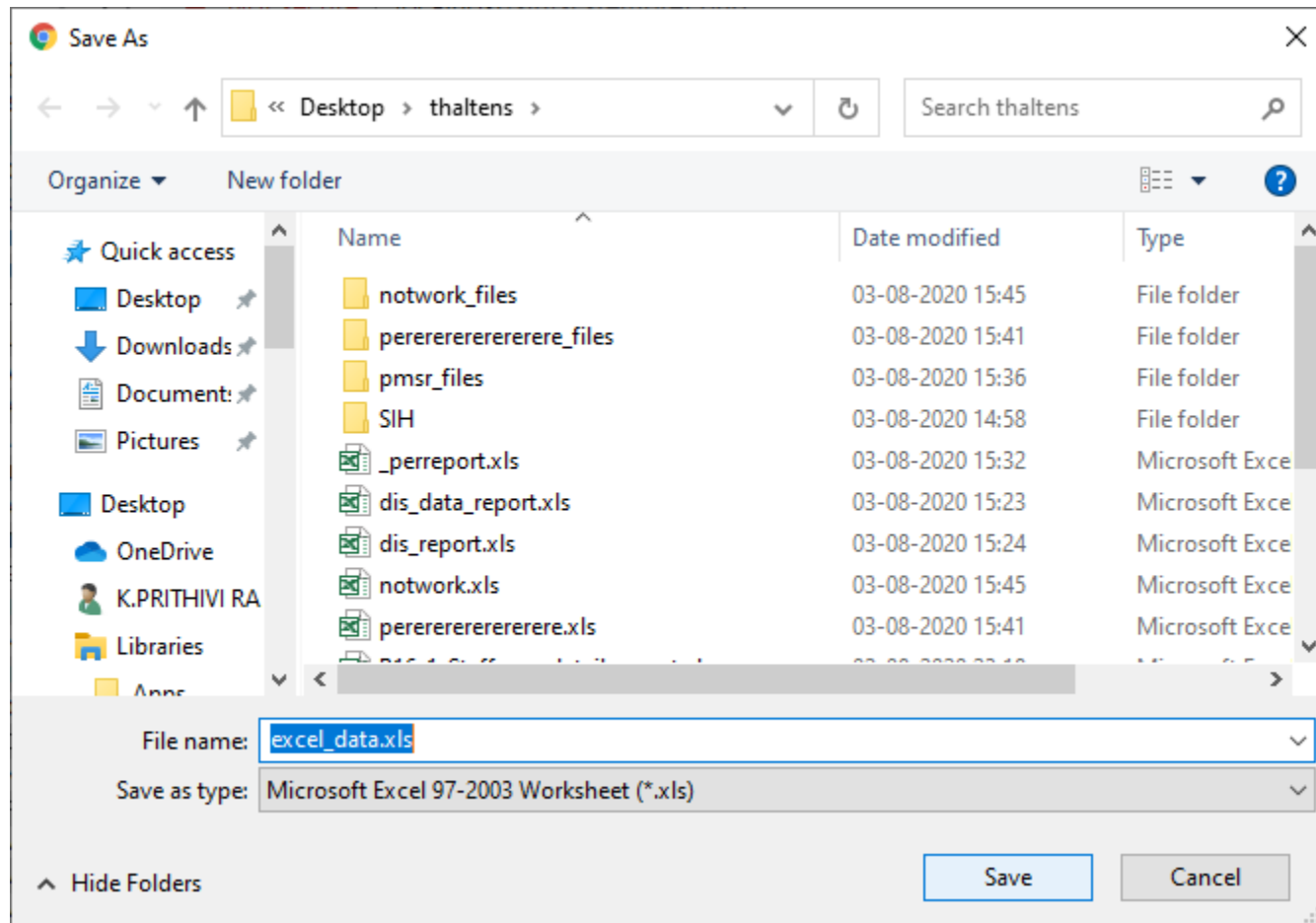
| District | HH issued jobcards SC | HH issued jobcards ST | HH issued jobcards Others | HH issued jobcards Total | HH Provided Employment SC | HH Provided Employment ST | HH Provided Employment Others | HH Provided Employment Total | HH Provided Employment Women | Persondays generated SC | Persondays generated ST | Persondays generated Total | Persondays generated Others | Persondays generated Women | Families Completed 100 Days SC | Families Completed 100 Days ST | Families Completed 100 Days Total | Families Completed 100 Days Others |
|-----------------|-----------------------|-----------------------|---------------------------|--------------------------|---------------------------|---------------------------|-------------------------------|------------------------------|------------------------------|-------------------------|-------------------------|----------------------------|-----------------------------|----------------------------|--------------------------------|--------------------------------|-----------------------------------|------------------------------------|
| ARIYALUR | 33300 | 1690 | 141258 | 176248 | 22638 | 1051 | 89683 | 113372 | 104926 | 247906 | 12020 | 963635 | 1223561 | 978565 | 0 | 0 | 0 | 0 |
| COIMBATORE | 49574 | 3913 | 106447 | 159934 | 19312 | 929 | 37341 | 57582 | 53954 | 400943 | 20700 | 817738 | 1239381 | 1109949 | 0 | 0 | 4 | 4 |
| CUDDALORE | 152576 | 1833 | 290683 | 445092 | 91422 | 848 | 154781 | 247051 | 230587 | 1313412 | 11162 | 1956905 | 3281479 | 2413533 | 6 | 0 | 6 | 12 |
| DHARMAPURI | 29198 | 5914 | 203947 | 239059 | 15559 | 3123 | 119881 | 138563 | 136660 | 176774 | 46494 | 1430162 | 1653430 | 1566326 | 0 | 0 | 0 | 0 |
| DINDIGUL | 79123 | 675 | 248156 | 327954 | 49733 | 246 | 158058 | 208037 | 194286 | 1012171 | 4494 | 3005990 | 4022655 | 3468320 | 0 | 0 | 8 | 8 |
| ERODE | 53599 | 2833 | 144317 | 200749 | 24282 | 1332 | 58648 | 84262 | 76754 | 588188 | 37130 | 1454622 | 2079940 | 1806982 | 0 | 0 | 4 | 4 |
| KANCHIPURAM | 126996 | 4433 | 205590 | 337019 | 72183 | 2236 | 110619 | 185038 | 170839 | 1122455 | 38323 | 1582517 | 2743295 | 2413906 | 0 | 0 | 0 | 0 |
| KANNIYAKUMARI | 5726 | 639 | 102767 | 109132 | 1863 | 77 | 22468 | 24408 | 21871 | 39337 | 1907 | 461604 | 502848 | 442913 | 0 | 0 | 5 | 5 |
| KARUR | 32501 | 86 | 115446 | 148033 | 21600 | 54 | 79335 | 100989 | 96435 | 533851 | 1219 | 1875842 | 2410912 | 2108265 | 6 | 0 | 28 | 34 |
| KRISHNAGIRI | 28365 | 1584 | 247444 | 277393 | 17807 | 876 | 158207 | 176890 | 169026 | 312073 | 14689 | 2603093 | 2929855 | 2709539 | 0 | 0 | 1 | 1 |
| MADURAI | 61499 | 495 | 230759 | 292753 | 40142 | 285 | 156191 | 196618 | 185156 | 560634 | 3133 | 2188289 | 2752056 | 2400897 | 1 | 0 | 0 | 1 |
| NAGAPATTINAM | 100737 | 346 | 151982 | 253065 | 85428 | 264 | 120059 | 205751 | 204274 | 2291419 | 6073 | 2741262 | 5038754 | 3735389 | 14 | 0 | 13 | 27 |
| NAMAKKAL | 60802 | 15039 | 153599 | 229440 | 33254 | 8419 | 78178 | 119851 | 113071 | 586403 | 157739 | 1428334 | 2172476 | 1946348 | 0 | 0 | 4 | 4 |
| PERAMBALUR | 36171 | 385 | 84892 | 121448 | 26478 | 239 | 58935 | 85652 | 82529 | 324593 | 3351 | 727746 | 1055690 | 890893 | 1 | 0 | 2 | 3 |
| PUDUKKOTTAI | 49743 | 98 | 244831 | 294672 | 36003 | 64 | 177623 | 213690 | 206084 | 540128 | 871 | 2589374 | 3130373 | 2772497 | 0 | 0 | 2 | 2 |
| RAMANATHAPURAM | 52255 | 47 | 159336 | 211638 | 38488 | 35 | 114439 | 152962 | 146718 | 926970 | 918 | 2638983 | 3566871 | 3098163 | 10 | 0 | 34 | 44 |
| SALEM | 86474 | 28039 | 316656 | 431169 | 19885 | 6539 | 91565 | 117989 | 111372 | 274363 | 133204 | 1393680 | 1801247 | 1578798 | 0 | 6 | 8 | 14 |
| SIVAGANGAI | 41642 | 50 | 163030 | 204722 | 28425 | 31 | 114025 | 142481 | 135694 | 720662 | 850 | 2828837 | 3550349 | 3175218 | 2 | 0 | 7 | 9 |
| THANJAVUR | 88145 | 206 | 216222 | 304573 | 64985 | 125 | 147595 | 212705 | 205810 | 1375831 | 3079 | 3031263 | 4410173 | 3681152 | 2 | 0 | 2 | 4 |
| THE NILGIRIS | 14132 | 3267 | 33796 | 51195 | 4716 | 1063 | 12398 | 18177 | 16129 | 103251 | 25187 | 241661 | 370099 | 305064 | 0 | 0 | 1 | 1 |
| THENI | 30898 | 170 | 103581 | 134649 | 9574 | 64 | 35570 | 45208 | 42366 | 102290 | 925 | 392409 | 495624 | 425672 | 0 | 0 | 2 | 2 |
| THOOTHUKKUDI | 57272 | 161 | 148987 | 206420 | 34624 | 88 | 81267 | 115979 | 108983 | 706306 | 1750 | 1710149 | 2418205 | 2133429 | 1 | 0 | 2 | 3 |
| TIRUCHIRAPPALLI | 70875 | 3019 | 250688 | 324582 | 53642 | 2398 | 188493 | 244533 | 232070 | 1025593 | 65284 | 3421360 | 4512337 | 3908779 | 0 | 0 | 1 | 1 |
| TIRUNELVELI | 81053 | 364 | 217901 | 299318 | 45636 | 168 | 111869 | 157673 | 148943 | 1002158 | 4452 | 2498515 | 3505125 | 3056461 | 1 | 0 | 1 | 2 |
| TIRUPPUR | 66426 | 234 | 137477 | 204137 | 34454 | 94 | 63571 | 98119 | 88805 | 924262 | 2652 | 1720509 | 2647423 | 2294794 | 0 | 0 | 9 | 9 |
| TIRUVALLUR | 105149 | 5801 | 185731 | 296681 | 61231 | 3513 | 113022 | 177766 | 167058 | 987470 | 60119 | 1779769 | 2827358 | 2604568 | 0 | 0 | 2 | 2 |
| TIRUVANNAMALAI | 107456 | 20261 | 396114 | 523831 | 84406 | 9248 | 313755 | 407409 | 371600 | 1527800 | 159926 | 5283541 | 6971267 | 5819514 | 0 | 0 | 3 | 3 |
| TIRUVARUR | 97711 | 116 | 113005 | 210832 | 77935 | 65 | 69829 | 147829 | 146146 | 2172736 | 1537 | 1699389 | 3873662 | 2905303 | 12 | 0 | 5 | 17 |
| VELLORE | 97460 | 8486 | 323281 | 429227 | 54114 | 4247 | 186579 | 244940 | 233612 | 979311 | 89800 | 3014522 | 4083633 | 3686847 | 0 | 0 | 0 | 0 |
| VILLUPURAM | 205793 | 15141 | 493484 | 714418 | 155630 | 8138 | 688118 | 988868 | 1611008 | 1000000 | 1000000 | 1000000 | 1000000 | 1000000 | 2 | 0 | 9 | 11 |
| VIRUDHUNAGAR | 63585 | 82 | 176867 | 240534 | 36463 | 26 | 100000 | 100000 | 100000 | 100000 | 100000 | 100000 | 100000 | 100000 | 0 | 0 | 8 | 8 |
| Total | 2166236 | 125407 | 6108274 | 8399917 | 1361912 | 55885 | 1000000 | 1000000 | 1000000 | 1000000 | 1000000 | 1000000 | 1000000 | 1000000 | 58 | 6 | 171 | 235 |


meet.google.com is sharing your screen.


Stop sharing

Hide

Activate Windows
Go to Settings to activate Windows.



 **The Mahatma Gandhi National Rural Employment Guarantee Act 2005**
Ministry of Rural Development
Government of India



SC ST Employment Status

District: COIMBATORE

HH issued jobcards

| | |
|--------|--------|
| SC | 233 |
| ST | 234 |
| Others | 234234 |
| Women | 232342 |

HH issued Employment

| | |
|--------|--------|
| SC | 234234 |
| ST | 335 |
| Others | 3453 |
| Total | 3434 |
| Women | 345345 |


No. of Persondays generated

| | |
|--------|--------|
| SC | 343 |
| ST | 3453 |
| Others | 333 |
| Total | 1231 |
| Women | 123123 |

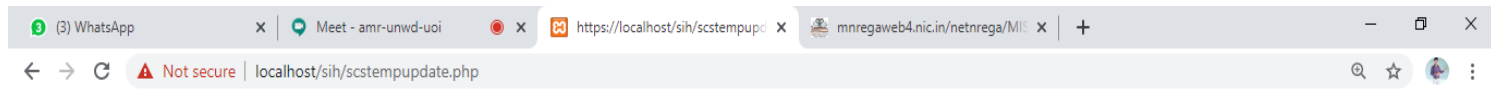
Families completed

| | |
|--------|-----|
| SC | 234 |
| ST | 234 |
| Others | 23 |
| Total | 23 |

UPDATE **RESET**



Activate Windows
Go to Settings to activate Windows.



Data Updated successfully

Go back


https://localhost/sih/scstempupdate.html


Activate Windows
Go to Settings to activate Windows.

Deleting records

(3) WhatsApp x SC/ST jobcards x mnregaweb4.nic.in/netnrega/MI x +

File | C:/xampp/htdocs/sih/scstempdelet.html

 **The Mahatma Gandhi National Rural Employment Guarantee Act 2005**
Ministry of Rural Development
Government of India



SC/ST Retrive records

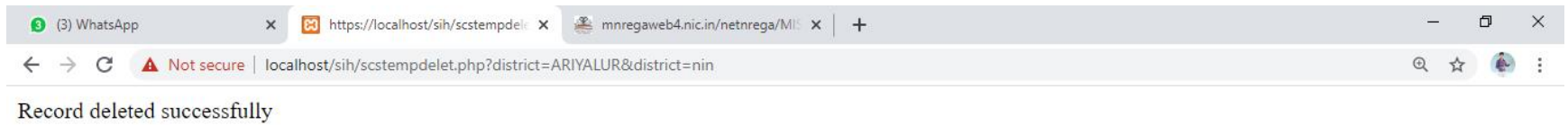
***Kindly select the District and academic year**

District

Academic Year

***By Clicking This button You are deleting the data**

Acknowledgement





The Mahatma Gandhi National Rural Employment Guarantee Act 2005
Ministry of Rural Development
Government of India





Progress report-Retrive

***Kindly select the District and academic year**

State

TAMILNADU ▼

District

ARIYALUR ▼

Block

ANDIMADAM ▼

Academic Year

2020-2021 ▼

GoBack

SEARCH

PROGRESS REPORT 2020-2021

| State | DistrictSC | Block | Total Jobcards Applied | Total jobcards issued | Total Jobcards SC | Total Jobcards ST | Total Jobcards non SC/ST | Total HH demanded | Total Persons Demanded | Total HH job allocated | Total person job Allocated |
|------------|------------|-------------|------------------------|-----------------------|-------------------|-------------------|--------------------------|-------------------|------------------------|------------------------|----------------------------|
| TAMIL NADU | ARIYALUR | ANDIMADAM | 30281 | 29512 | 5604 | 463 | 23445 | 26804 | 40707 | 26803 | 40705 |
| TAMIL NADU | ARIYALUR | ARIYALUR | 32862 | 32389 | 6200 | 212 | 25977 | 29389 | 43105 | 29388 | 43104 |
| TAMIL NADU | ARIYALUR | JAYAMKONDAM | 33226 | 32937 | 8403 | 683 | 23851 | 31383 | 56231 | 31383 | 56231 |
| TAMIL NADU | ARIYALUR | SENDURAI | 40195 | 38987 | 7706 | 176 | 31105 | 34602 | 50643 | 34602 | 50643 |
| TAMIL NADU | ARIYALUR | T.PALUR | 27804 | 26500 | 5448 | 406 | 20646 | 24221 | 36960 | 24221 | 36959 |
| TAMIL NADU | ARIYALUR | THIRUMANUR | 29508 | 28963 | 6104 | 9 | 22850 | 26765 | 39185 | 26765 | 39184 |
| TAMILNADU | ARIYALUR | ANDIMADAM | 123 | 123123 | 123123 | 121 | 12112 | 12123 | 3332 | 3422 | 232323 |
| TAMILNADU | ARIYALUR | ANDIMADAM | 876567 | 7656 | 8687687 | 687 | 878 | 8687 | 8668 | 876 | 876 |

Export Table Data To Excel File

smart india hackathon © 2020

meet.google.com is sharing your screen. Stop sharing Hide

Activate Windows
Go to Settings to activate Windows.

(3) WhatsApp x PHYSICAL PROGRESS x mnregaweb4.nic.in/netnrega/M... x +

File | C:/xampp/htdocs/sih/progrepmain.html

सत्यमेव जयते

GOVERNMENT OF INDIA

PROGRESS REPORT

You can Insert, Update and Retrieve the Data!!!

INSERT

RETRIVE DATA

UPDATE

DELETE DATA

Go To HomePage

Check the report click Counts!

You have clicked the button 27 time(s).

MAHATMA GANDHI

SANTABANTA.COM

Activate Windows
Go to Settings to activate Windows.

Smart India Hackathon@2020

GRAPHICAL REPRESENTATION OF THE DATA FOR DATA ANALYSIS

```
dataset = pandas_csv(csv(user_input))
```

Enter the path of your CSV file:

Enter the path of your CSV file:

StateName
Totalhouseholdsappliedforjobcard
TotaljobcardsSC
TotaljobcardsST
TotaljobcardsnonSCST
Totalpersonsallottedwork
Totalpersonsworked
TotalpersondaysworkednonSCST
Totalperson-daysworkedbySCs
TotalpersondaysworkedbySTs
TotalSCHouseholdsover100daylimit
TotalSTHouseholdsover100daylimit

Kindly Select the Columns Which you want to plot
enter x Axis

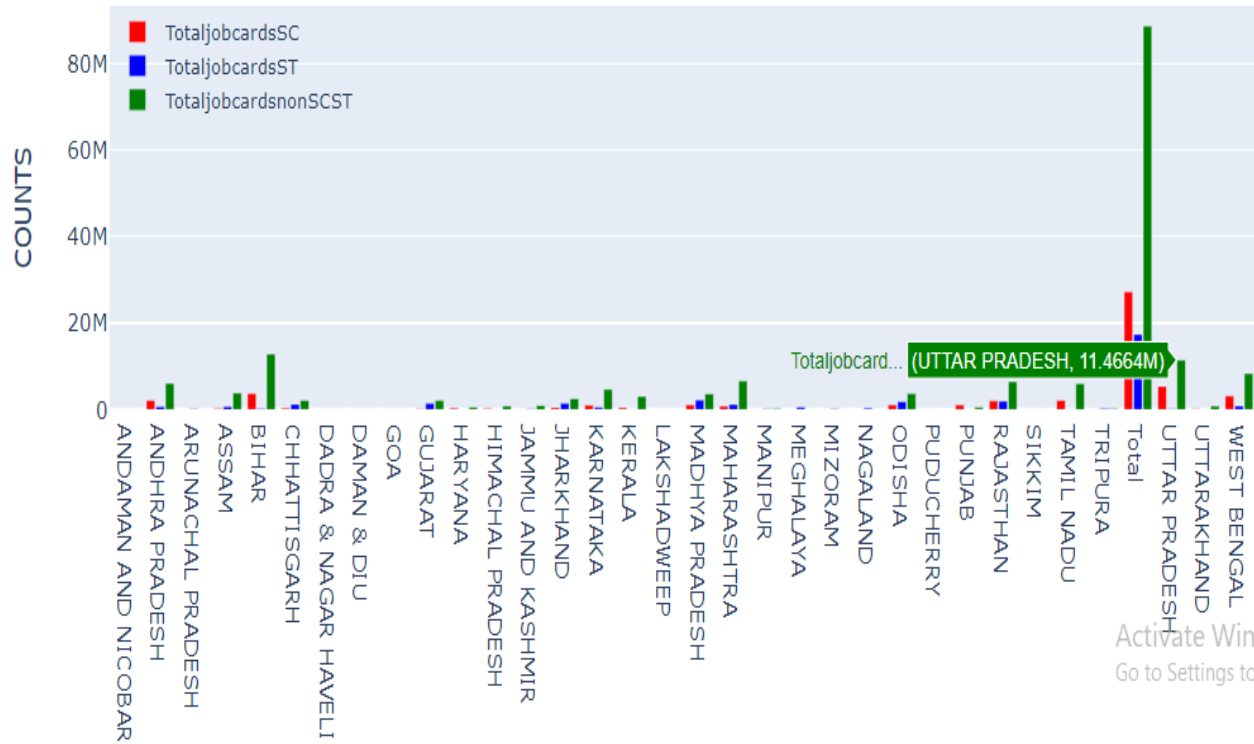
StateName

Kindly Select the Columns Which you want to plot
enter x Axis
StateName

Enter Y Axis separated by comma jobcardsSC,TotaljobcardsST,TotaljobcardsnonSCST

In []:

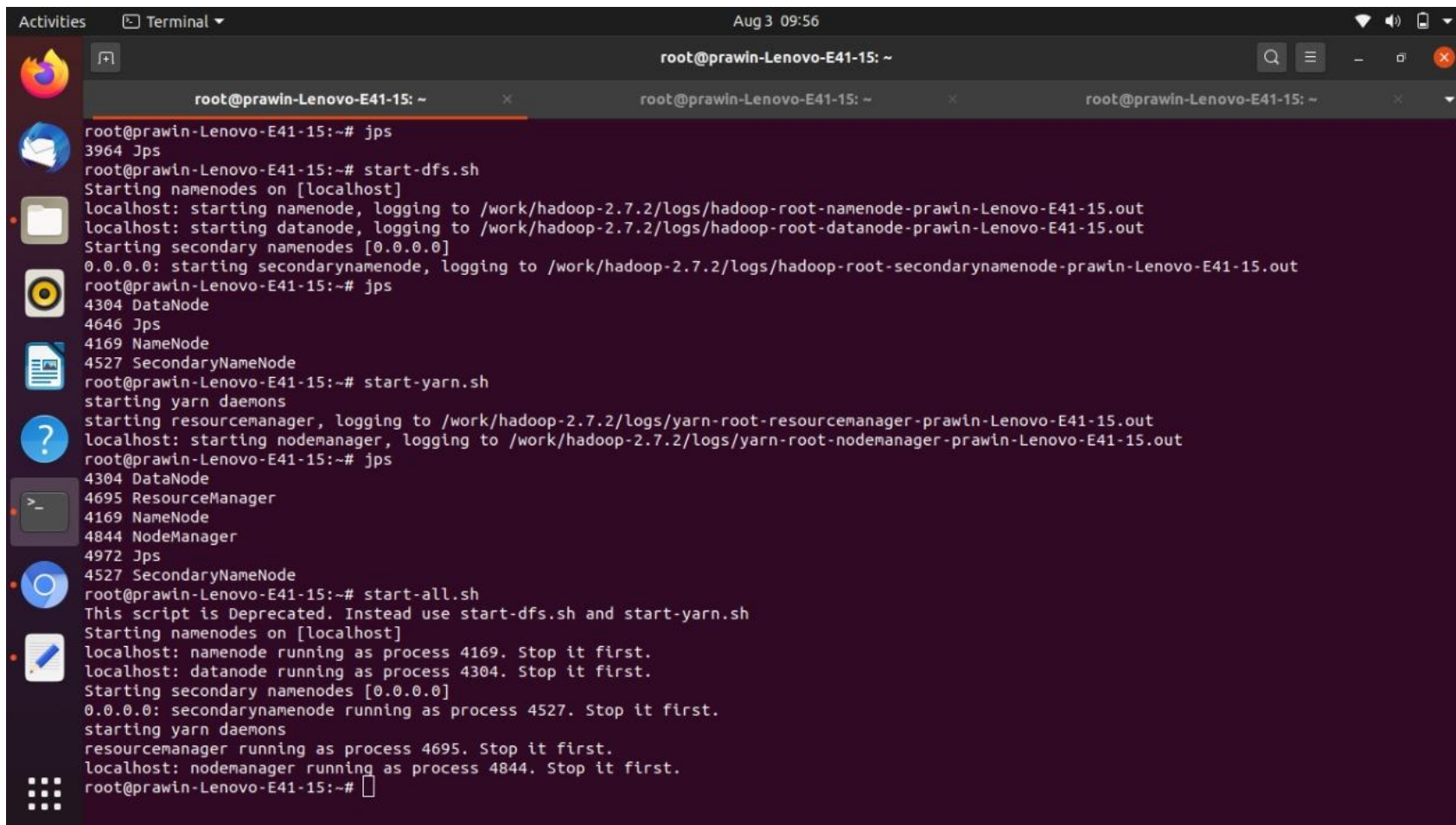
REPORT



Activate Windows
Go to Settings to activate Windows.

Spark streaming

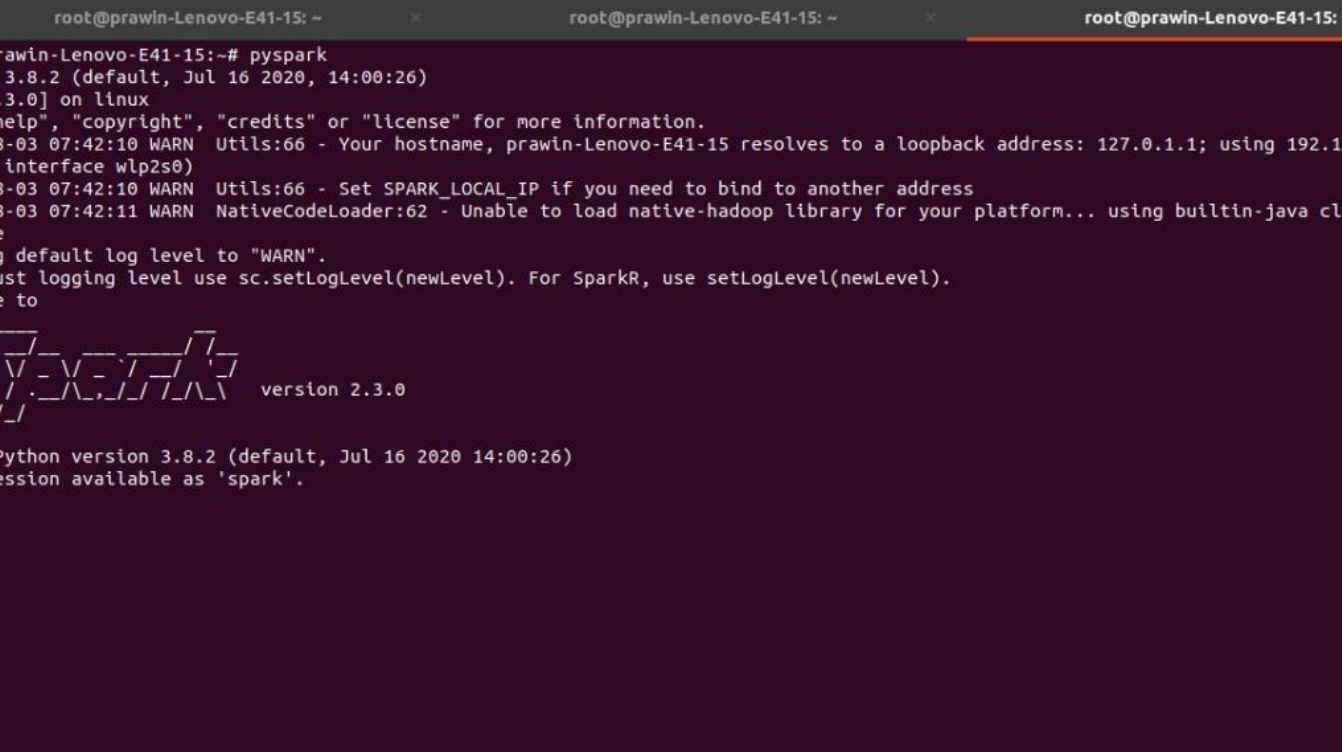
Starting Daemons services



```
Activities  Terminal  Aug 3 09:56
root@prawin-Lenovo-E41-15: ~

root@prawin-Lenovo-E41-15:~# jps
3964 Jps
root@prawin-Lenovo-E41-15:~# start-dfs.sh
Starting namenodes on [localhost]
localhost: starting namenode, logging to /work/hadoop-2.7.2/logs/hadoop-root-namenode-prawin-Lenovo-E41-15.out
localhost: starting datanode, logging to /work/hadoop-2.7.2/logs/hadoop-root-datanode-prawin-Lenovo-E41-15.out
Starting secondary namenodes [0.0.0.0]
0.0.0.0: starting secondarynamenode, logging to /work/hadoop-2.7.2/logs/hadoop-root-secondarynamenode-prawin-Lenovo-E41-15.out
root@prawin-Lenovo-E41-15:~# jps
4304 DataNode
4646 Jps
4169 NameNode
4527 SecondaryNameNode
root@prawin-Lenovo-E41-15:~# start-yarn.sh
starting yarn daemons
starting resourcemanager, logging to /work/hadoop-2.7.2/logs/yarn-root-resourcemanager-prawin-Lenovo-E41-15.out
localhost: starting nodemanager, logging to /work/hadoop-2.7.2/logs/yarn-root-nodemanager-prawin-Lenovo-E41-15.out
root@prawin-Lenovo-E41-15:~# jps
4304 DataNode
4695 ResourceManager
4169 NameNode
4844 NodeManager
4972 Jps
4527 SecondaryNameNode
root@prawin-Lenovo-E41-15:~# start-all.sh
This script is Deprecated. Instead use start-dfs.sh and start-yarn.sh
Starting namenodes on [localhost]
localhost: namenode running as process 4169. Stop it first.
localhost: datanode running as process 4304. Stop it first.
Starting secondary namenodes [0.0.0.0]
0.0.0.0: secondarynamenode running as process 4527. Stop it first.
starting yarn daemons
resourcemanager running as process 4695. Stop it first.
localhost: nodemanager running as process 4844. Stop it first.
root@prawin-Lenovo-E41-15:~#
```

PySpark



```
root@prawin-Lenovo-E41-15: ~# pyspark
Python 3.8.2 (default, Jul 16 2020, 14:00:26)
[GCC 9.3.0] on linux
Type "help", "copyright", "credits" or "license" for more information.
2020-08-03 07:42:10 WARN Utils:66 - Your hostname, prawin-Lenovo-E41-15 resolves to a loopback address: 127.0.1.1; using 192.168.43.166 instead (on interface wlp2s0)
2020-08-03 07:42:10 WARN Utils:66 - Set SPARK_LOCAL_IP if you need to bind to another address
2020-08-03 07:42:11 WARN NativeCodeLoader:62 - Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
Welcome to

  ____      _
 / ___|  __| | | |
 \___ \  | | | | | |
  ___) | | | | | | |
 |_____|_|_|_|_|_|_|

 version 2.3.0

Using Python version 3.8.2 (default, Jul 16 2020 14:00:26)
SparkSession available as 'spark'.
>>> 
```

Map Reducing method

```
Activities Terminal Sat Aug 1 11:00 AM
prawin@prawin-Lenovo-E41-15: ~
20/08/01 11:00:30 INFO ShuffleBlockFetcherIterator: Getting 1 (539.0 B) non-empty blocks including 1 (539.0 B) local and 0 (0.0 B) host-local and 0 (0.0 B) remote blocks
20/08/01 11:00:30 INFO ShuffleBlockFetcherIterator: Started 0 remote fetches
20/08/01 11:00:30 INFO PythonRunner: Times: total = 17, boot = -9932, in
20/08/01 11:00:30 INFO PythonRunner: Times: total = 44, boot = -40, init
20/08/01 11:00:30 INFO Executor: Finished task 0.0 in stage 1406.0 (TID
20/08/01 11:00:30 INFO TaskSetManager: Finished task 0.0 in stage 1406.0
20/08/01 11:00:30 INFO TaskSchedulerImpl: Removed TaskSet 1406.0, whose
20/08/01 11:00:30 INFO DAGScheduler: ResultStage 1406 (runJob at PythonR
20/08/01 11:00:30 INFO DAGScheduler: Job 703 is finished. Cancelling pot
20/08/01 11:00:30 INFO TaskSchedulerImpl: Killing all running tasks in s
20/08/01 11:00:30 INFO DAGScheduler: Job 703 finished: runJob at PythonR
-----
Time: 2020-08-01 11:00:30
-----
('MGNREGA', 1)
('houses', 1)
('large', 2)
('of', 8)
('country', 1)
('50', 1)
('consisting', 2)
('year', 2)
('starting', 1)
('primarily', 1)
...
20/08/01 11:00:30 INFO JobScheduler: Finished job streaming job 1596259830000 ms.0 from job set of time 1596259830000 ms
20/08/01 11:00:30 INFO JobScheduler: Total delay: 0.324 s for time 1596259830000 ms (execution: 0.263 s)
20/08/01 11:00:30 INFO PythonRDD: Removing RDD 2806 from persistence list
20/08/01 11:00:30 INFO BlockRDD: Removing RDD 2801 from persistence list
20/08/01 11:00:30 INFO SocketInputDStream: Removing blocks of RDD BlockRDD[2801] at socketTextStream at NativeMethodAccessorImpl.java:0 of time 1596259830000 ms
20/08/01 11:00:30 INFO ReceivedBlockTracker: Deleting batches: 1596259810000 ms
20/08/01 11:00:30 INFO InputInfoTracker: remove old batch metadata: 1596259810000 ms
20/08/01 11:00:30 INFO BlockManager: Removing RDD 2806
20/08/01 11:00:30 INFO BlockManager: Removing RDD 2801
```

Map Reducing Task

```
Activities Terminal Sat Aug 1 10:58 AM
prawin@prawin-Lenovo-E41-15: ~
20/08/01 10:58:40 INFO ShuffleBlockFetcherIterator: Getting 1 (80.0 B) non-empty blocks including 1 (80.0 B) local and 0 (0.0 B) host-local and 0 (0.0 B) remote blocks
20/08/01 10:58:40 INFO ShuffleBlockFetcherIterator: Started 0 remote fetches
20/08/01 10:58:40 INFO BlockManagerInfo: Removed broadcast_685_piece0 on
20/08/01 10:58:40 INFO PythonRunner: Times: total = 16, boot = -70, init
20/08/01 10:58:40 INFO PythonRunner: Times: total = 62, boot = -94, init
20/08/01 10:58:40 INFO Executor: Finished task 0.0 in stage 1364.0 (TID
20/08/01 10:58:40 INFO TaskSetManager: Finished task 0.0 in stage 1364.0
20/08/01 10:58:40 INFO TaskSchedulerImpl: Removed TaskSet 1364.0, whose
20/08/01 10:58:40 INFO DAGScheduler: ResultStage 1364 (runJob at PythonR
20/08/01 10:58:40 INFO DAGScheduler: Job 682 is finished. Cancelling pot
20/08/01 10:58:40 INFO TaskSchedulerImpl: Killing all running tasks in s
20/08/01 10:58:40 INFO DAGScheduler: Job 682 finished: runJob at PythonR
-----
Time: 2020-08-01 10:58:40
-----
('hakathon', 1)
('india', 1)
-----
20/08/01 10:58:40 INFO JobScheduler: Finished job streaming job 15962597
20/08/01 10:58:40 INFO JobScheduler: Total delay: 0.505 s for time 15962
20/08/01 10:58:40 INFO PythonRDD: Removing RDD 2718 from persistence lis
20/08/01 10:58:40 INFO BlockRDD: Removing RDD 2713 from persistence list
20/08/01 10:58:40 INFO SocketInputDStream: Removing blocks of RDD BlockR
9720000 ms
20/08/01 10:58:40 INFO ReceivedBlockTracker: Deleting batches: 1596259700000 ms
20/08/01 10:58:40 INFO InputInfoTracker: remove old batch metadata: 1596259700000 ms
20/08/01 10:58:40 INFO BlockManager: Removing RDD 2718
20/08/01 10:58:40 INFO BlockManager: Removing RDD 2713
20/08/01 10:58:40 INFO BlockManagerInfo: Removed input-0-1596259706800 on 192.168.43.253:46687 in memory (size: 16.0 B, free: 434.4 MiB)
20/08/01 10:58:40 INFO BlockManagerInfo: Removed input-0-1596259708000 on 192.168.43.253:46687 in memory (size: 10.0 B, free: 434.4 MiB)
20/08/01 10:58:42 INFO MemoryStore: Block input-0-1596259721800 stored as values in memory (estimated size 17.0 B, free 434.3 MiB)
20/08/01 10:58:42 INFO BlockManagerInfo: Added input-0-1596259721800 in memory on 192.168.43.253:46687 (size: 17.0 B, free: 434.4 MiB)
20/08/01 10:58:42 WARN RandomBlockReplicationPolicy: Expecting 1 replicas with only 0 peer/s.
20/08/01 10:58:42 INFO BlockManagerInfo: Removed broadcast_687_piece0 on 192.168.43.253:46687 in memory (size: 5.2 KiB, free: 434.4 MiB)
20/08/01 10:58:42 WARN BlockManager: Block input-0-1596259721800 replicated to only 0 peer(s) instead of 1 peers
20/08/01 10:58:42 INFO BlockGenerator: Pushed block input-0-1596259721800

```

Hadoop Web UI

Activities Chromium Web Browser Aug 3 13:08

Namenode information - Chromium

Namenode informa x PySparkShell - Spark x Naive Bayes - MLlib x Multi-Class Text Cla x M-Vigneshwaran-gi x (9) WhatsApp x +

← → ↻ ⌂ ⓘ Not secure | 192.168.43.166:50070/dfshealth.html#tab-overview ☆ 🗨️ 🍌 ⬇️ 🐾 🌐 👤 ⋮

Hadoop Overview Datanodes Datanode Volume Failures Snapshot Startup Progress Utilities ▾

Overview 'localhost:9000' (active)

| | |
|----------------|---|
| Started: | Mon Aug 03 08:30:50 IST 2020 |
| Version: | 2.7.2, rb165c4fe8a74265c792ce23f546c64604acf0e41 |
| Compiled: | 2016-01-26T00:08Z by jenkins from (detached from b165c4f) |
| Cluster ID: | CID-61eca170-6f5e-44e6-a8cb-e746cae2d35c |
| Block Pool ID: | BP-401175755-127.0.1.1-1596418513830 |

Summary

Security is off.

Safemode is off.

36 files and directories, 12 blocks = 48 total filesystem object(s).

Heap Memory used 79.38 MB of 162 MB Heap Memory. Max Heap Memory is 889 MB.

Non Heap Memory used 61.21 MB of 62.22 MB Committed Non Heap Memory. Max Non Heap Memory is -1 B.

Moving data from Local file to HDFS.

Activities Chromium Web Browser Aug 3 10:00

Browsing HDFS - Chromium

Browsing HDFS (8) WhatsApp

Not secure | 192.168.43.166:50070/explorer.html#/demo/day

Hadoop Overview Datanodes Snapshot Startup Progress Utilities

Browse Directory

/demo/day Go!

| Permission | Owner | Group | Size | Last Modified | Replication | Block Size | Name |
|------------|-------|------------|---------|----------------------|-------------|------------|------------------------------|
| -rw-r--r-- | root | supergroup | 1.88 KB | 03/08/2020, 08:05:45 | 1 | 128 MB | data_set.csv |
| -rw-r--r-- | root | supergroup | 445 B | 03/08/2020, 09:18:10 | 1 | 128 MB | demodata.csv |

Hadoop, 2015.

Activities

Chromium Web Browser

Aug 3 13:14

Browsing HDFS - Chromium

Browsing HDFS

PySparkShell - Spark

Naive Bayes - MLlib

Multi-Class Text Cla

M-Vigneshwaran-gi

(9) WhatsApp

Not secure | 192.168.43.166:50070/explorer.html#/demo/day

Hadoop

Overview

Datanodes

Snapshot

Startup Progress

Utilities

Browse Directory

/demo/day

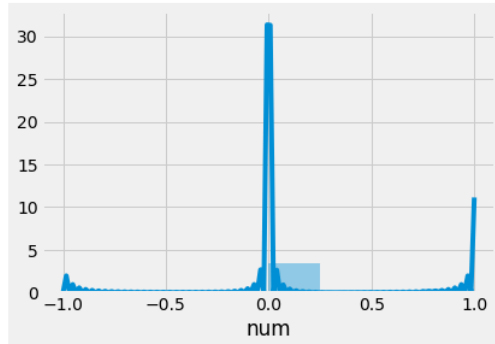
Go!

| Permission | Owner | Group | Size | Last Modified | Replication | Block Size | Name |
|------------|-------|------------|---------|----------------------|-------------|------------|------------------------------------|
| -rw-r--r-- | root | supergroup | 1.88 KB | 03/08/2020, 08:05:45 | 1 | 128 MB | data_set.csv |
| -rw-r--r-- | root | supergroup | 445 B | 03/08/2020, 09:18:10 | 1 | 128 MB | demodata.csv |
| drwxr-xr-x | root | supergroup | 0 B | 03/08/2020, 10:49:35 | 0 | 0 B | processed_data.csv |
| -rw-r--r-- | root | supergroup | 354 B | 03/08/2020, 10:13:13 | 1 | 128 MB | std.csv |

Hadoop, 2015.

Prediction algorithm

Out[16]: <matplotlib.axes._subplots.AxesSubplot at 0x7feb79b972b0>



```
In [17]: print(metrics.mean_absolute_error(y_test,predictions))
```

0.14473684210526316

```
In [18]: print(metrics.mean_squared_error(y_test,predictions))
```

0.14473684210526316

```
In [19]: print(np.sqrt(metrics.mean_squared_error(y_test,predictions)))
```

0.3804429551263411

```
In [20]: from sklearn.metrics import classification_report
```

```
In [21]: print(classification_report(y_test, predictions))
```

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0 | 0.81 | 0.92 | 0.86 | 38 |
| 1 | 0.91 | 0.79 | 0.85 | 38 |
| micro avg | 0.86 | 0.86 | 0.86 | 76 |
| macro avg | 0.86 | 0.86 | 0.85 | 76 |
| weighted avg | 0.86 | 0.86 | 0.85 | 76 |