



DATA ANALYTICS DONE BY REAL TIME DATA USING HADOOP AND SPARK

¹Bindu Alugula

¹M.tech, Department of Computer Science,
Andhra University, Visakhapatnam, AP, India

Abstract : Data Analytics Done by Real Time Data using Hadoop and Spark is a Business Analytics Application. Its main goal is to migrate and transfer the data from client and Application server to HIVE with the help of Spark. This Application is generally handled by Data Engineer. Develop the data mindset and analytical skills to make informed business decisions here the main decision which I made is deals with data which can be migrated and transfer using Avro format and finally comes in parquet format which is default format of spark is dump in hive. Generally we have Client Server, Application Server, Spark and Hive. My Project Deals with these software's only. In Client Server data is stored in form of Key Value pair and Maximum deals with mysql. In Application Server is a Key Data. Spark is used for the speed access and manipulate of the data with the help of .jar files .Spark make use of HBASE to convert Avro format to data frames .These Data Frames are dump into HIVE in tabular format (parquet). Benefit is Client can monitor the transfer through log files. Finally internal data serialization and data migration is done.

KEYWORDS:

Hadoop, Real Time, Data Engineer, Business Analytics, Key Value, Avro Format, Spark, Hive, HBase, Application Server, Client Server, MySQL.

INTRODUCTION:

PURPOSE

In the current scenario, migrating and transferring data is very important .This is very easy if our data is of small size .What if it's a large size of data which we want to transfer from one place to other And in this pandemic time my code works real good I have done data engineering part which transfers data from avro format to data frame with the use of Spark, HBase and Hive.

SCOPE

The Scope of this application is it can be monitor by client through log files and operations and data transfer is done by client and application server with access of the .jar file and code of the data engineer. The Large data is transfer from one system to other.

EXISTING SYSTEM

These days everything is deal with data only .If it is small data transfer it is easy but the data size is increased every information is important these days. In older days data use to be used around in Megabytes But Now a Days Data is estimated to be in ZettaBytes So Transfer of data from one to other places takes time if we so old software's.

Disadvantages

- Data transmission is slow and loss of data packets can be seen.
- Bad in Optimizing & Improving the Customer Experience
- There is no personalization & less Services

Advantages

- Cost optimization
- Improve efficiency
- Foster competitive pricing.
- Boost sales and retain customer loyalty
- Innovate
- Focus on the local environment.
- Control and monitor online reputation

PROJECT DESCRIPTION

The project “**DATA ANALYTICS DONE BY REAL TIME DATA USING HADOOP AND SPARK**” is developed based on information related to clients who want to collect the data from old machine and want to transfer to new systems .This project mainly includes the maintenance of client and application server where the key value pair data is present and this is transferred to data frames into HIVE tabular format with the help of spark. Usually Spark is engineered from the bottom-up for performance. Spark can be 100 xs faster than Hadoop for large scale data processing by exploiting in memory computing and other optimizations. We have done in Spark for easy to use APIs for Operating on Large datasets. It is Unified Engine that why is transfer the data better than before existing systems.

First we have to start terminal Go to Start and press terminal in Ubuntu. Next we have to set Server with the command start – yarn.sh, start –dfs.sh, jps, start –hbase.sh or we can give like start –all.sh.

Second we have to give jar files to spark with the command spark-shell --jars /home/hadoop/hbasesparkjars/hbase-protocol-2.1.0.jar,/home/hadoop/hbasesparkjars/hbase-client-2.1.0.jar,/home/hadoop/hbasesparkjars/hbase-common-2.1.0.jar,/home/hadoop/hbasesparkjars/hbase-protocol-shaded-2.1.0.jar,/home/hadoop/hbasesparkjars/hbase-server-2.1.0.jar,/home/hadoop/hbasesparkjars/hbase-shaded-miscellaneous-2.1.0.jar,/home/hadoop/hbasesparkjars/hbase-shaded-netty-2.1.0.jar,/home/hadoop/hbasesparkjars/hbase-shaded-protobuf-2.1.0.jar,/home/hadoop/hbasesparkjars/hbase-spark-2.0.0-alpha4.jar,/home/hadoop/hbasesparkjars/htrace-core4-4.2.0-incubating.jar,/home/hadoop/hbasesparkjars/shc-core-1.1.1-2.1-s_2.11.jar

we have launch the spark and open with supporting jars

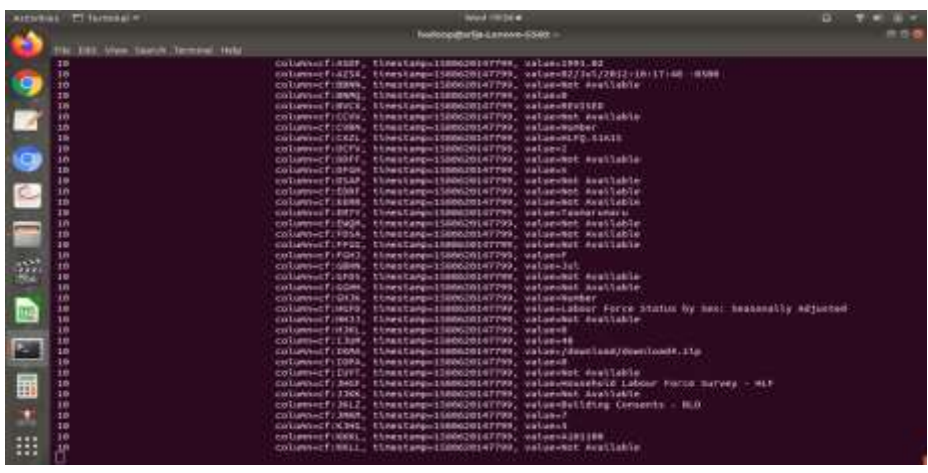
Third we import the classes like org.apache.spark.sql.{SQLContext,_},spark.sql.execution.datasources.hbase._,org.apache.spark.{SparkConf,SparkContext},sqlContext.implicits._,org.apache.hadoop.hbase.protobuf.generated.MasterProtos._,org.apache.hadoop.hbase.protobuf._,RequestConverter

Fourth, creating Avro format for retrieve data from HBase for that created method

Fifth, Reading HBase and SQL with spark and then AVRO format will be created then retrieve data from HBase and finally transformed data and dumping from HBase to hive.

RESULTS:

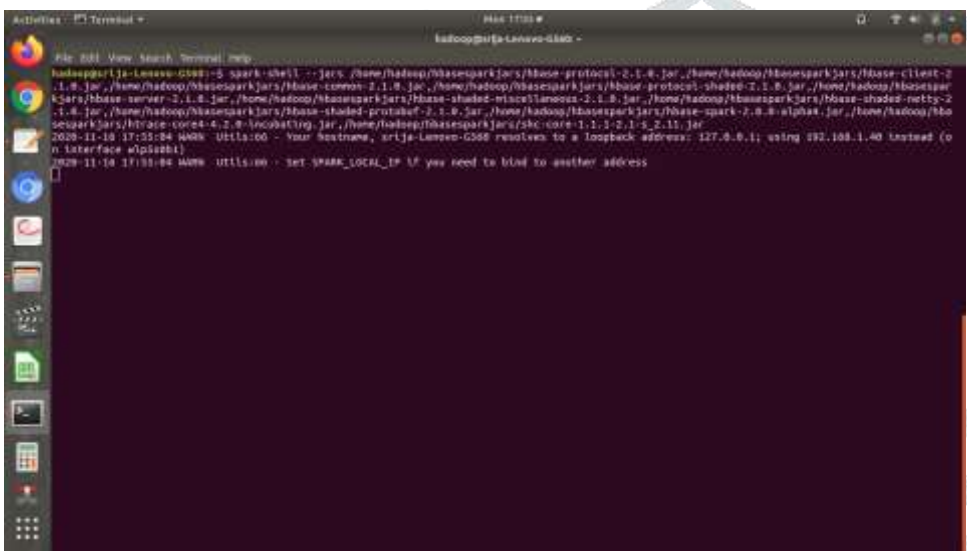
Data Tables



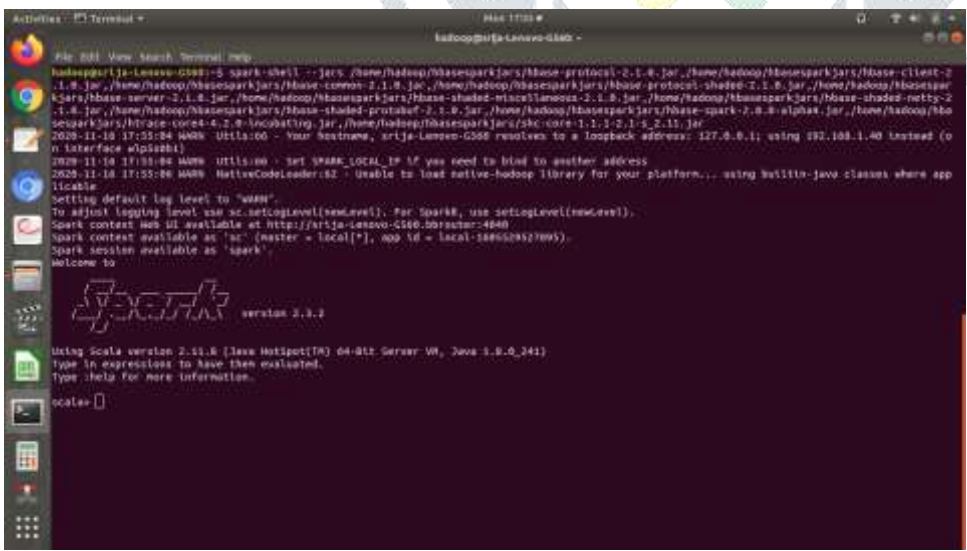
STEP 2 RUNNING SPARKS WITH JARS



STEPS 2.1 SPARK LAUNCHING



STEP 2.2 SPARK OPENED WITH SUPPORTING JARS



STEP 3 IMPORTING CLASSES

```
scala> import org.apache.spark.sql.{SQContext, _}
import org.apache.spark.sql._

scala> import org.apache.spark.sql.execution.datasources.hbase._
import org.apache.spark.sql.execution.datasources.hbase._

scala> import org.apache.spark.{SparkConf, SparkContext}
import org.apache.spark.{SparkConf, SparkContext}

scala> import sqContext.Implicits._
<console>:30: error: not found: value sqContext
import sqContext.Implicits._
        ^

scala> import org.apache.hadoop.hbase.protobuf.generated.MasterProtos._
<console>:31: error: object MasterProtos is not a member of package org.apache.hadoop.hbase.protobuf.generated
import org.apache.hadoop.hbase.protobuf.generated.MasterProtos._
        ^

scala> import org.apache.hadoop.hbase.protobuf._
import org.apache.hadoop.hbase.protobuf._

warning: Class org.apache.yetus.audience.InterfaceAudience not found - continuing with a stub.
warning: Class org.apache.yetus.audience.InterfaceAudience not found - continuing with a stub.
warning: Class org.apache.yetus.audience.InterfaceAudience not found - continuing with a stub.
warning: Class org.apache.yetus.audience.InterfaceAudience not found - continuing with a stub.
warning: Class org.apache.yetus.audience.InterfaceAudience not found - continuing with a stub.
warning: Class org.apache.yetus.audience.InterfaceAudience not found - continuing with a stub.
warning: Class org.apache.yetus.audience.InterfaceAudience not found - continuing with a stub.

scala> import org.apache.hadoop.hbase.protobuf.RequestCreator._
<console>:33: error: object RequestCreator is not a member of package org.apache.hadoop.hbase.protobuf
import org.apache.hadoop.hbase.protobuf.RequestCreator._
        ^

scala> import org.apache.hadoop.hbase.protobuf._
```

STEP 4 CREATING AVRO FORMAT FOR RETRIEVE DATA FROM HBASE FOR THAT CREATED METHOD.

```
scala> def columndata(columnname:Array[String]):String={var i = 1
    |   var camel = ""
    |   var stringdata = ""
    |   foreachColumn => {
    |     if(i != columnname.size)
    |     {
    |       stringdata = stringdata + "${eachElement.toUpperCase(i)}" + "(" + "${eachElement.toUpperCase(i)}" + ")" + "," + "type":"string"
    |       i = i + 1
    |     }
    |     else
    |     {
    |       stringdata = stringdata + "${eachElement.toUpperCase(i)}" + "(" + "${eachElement.toUpperCase(i)}" + ")" + "," + "type":"string"
    |     }
    |   }
    |   }
    |   }

scala> val finaldata = s"""
|   {
|     "table":{"namespace":"default", "name":"hbase_traction"},
|     "rowkey":"key",
|     "columns":{
|       "rowkey":{"cf":"rowkey", "col":"key", "type":"string"},
|       ${stringdata}
|     }
|   }
| }
| """

scala> columndata (columnname: Array[String])String
scala> []
```

STEP 5 READING HBASE AND SQL WITH SPARK

```
scala> def withCatalog(empcat:String)=
  |   spark.read.options(Map("hiveCatalog"->empcat)).
  |   format("org.apache.spark.sql.execution.datasources.hbase").load()
  |   withCatalog: (empcat: String)org.apache.spark.sql.DataSource

scala> val columnname = spark.read.format("orc").load("homed/hadoop/Downloads/JATIAS_Bases-Csv").selectCol("C1", "C2", "C3").rdd.map
  |   (_.asInstanceOf[os.type.String]).collect().toSeq

Loading class `com.mysql.jdbc.Driver`. This is deprecated. The new driver class is `com.mysql.cj.jdbc.Driver`. The driver is automatically registered via the SPI and manual loading of the driver class is generally unnecessary.

scala> []
```



```

Terminal
hadoop@qija-Linux-Virtual:
File Edit View Search Terminal Help
|-----|
| Val column = 47, columns Filter(=) Star(Star) Mac(Home) |
| df.select(column.head,column.tail(*)).show() |
|-----|
[20000_3g] [20000_507] [20000_508] [20000_2342] [20000_796] [20000_6969] [20000_232]
-----
| 2 | | | 1 | Not Available | Not Available | beautifulbear | A |
| 1 | | | 1 | Not Available | Not Available | beautifulbear | A |
| 2 | | | 1 | Not Available | Not Available | beautifulbear | A |
| 0 | | | 4 | Not Available | Not Available | beautifulbear | A |
| 0 | | | 2 | Not Available | Not Available | beautifulbear | A |
| 4 | Not Available | Not Available | Not Available | Mozilla/4.0 (comp... | 4037 |
| 0 | | | 4 | Not Available | Not Available | Not Available | Jakarta Commons-H... |
| 4 | Not Available | Not Available | Not Available | Mozilla/4.0 (comp... |
| 4 | Not Available | Not Available | Not Available | Mozilla/4.0 (comp... |
| 0 | | | 4 | Not Available | Not Available | Not Available | Mozilla/4.0 (comp... |
| 4 | Not Available | Not Available | Not Available | Mozilla/4.0 (comp... |
| 4 | Not Available | Not Available | Not Available | Mozilla/4.0 (comp... |
| 4 | Not Available | Not Available | Not Available | Mozilla/4.0 (comp... |
| 1 | | | 4 | Not Available | Not Available | Not Available | Jakarta Commons-H... |
| 0 | | | 2 | Not Available | Not Available | beautifulbear | 4037 |
| 4 | Not Available | Not Available | Not Available | Mozilla/4.0 (comp... |
| 1 | | | 4 | Not Available | Not Available | Not Available | Mozilla/4.0 (comp... |
| 1 | | | 4 | Not Available | Not Available | Not Available | Mozilla/4.0 (comp... |
| 0 | | | 4 | Not Available | Not Available | Not Available | Mozilla/4.0 (comp... |
|-----|
only showing top 20 rows
|-----|
[20000_55]
-----
| Number |
| Number |
| Number |
| SQM |
| Number |
|-----|

```

```

Terminal
hadoop@qija-Linux-Virtual:
File Edit View Search Terminal Help
|-----|
| SQM |
|-----|
|-----|
only showing top 20 rows
|-----|
[20000_333] [20000_55] [20000_dfg] [20000_6] [20000_654] [20000_dfg] [20000_wert222]
-----
| 1990-04-01 | Number | 9209 | Not Available | Male | /products | 130 |
| 1990-04-01 | Number | 9209 | Not Available | Male | /demo | 110 |
| 1990-04-01 | Number | 9209 | Not Available | Male | /download/download... | 33 |
| 1990-04-01 | Number | 9209 | Not Available | Male | /feeds/press | 0 |
| 1990-04-01 | SQM | 9274 | Not Available | Total Both Sexes | /demo | 0 |
| 1990-08-01 | Number | 4977 | Not Available | Total Both Sexes | /product/products | 0 |
| 1992-12-01 | SQM | 117 | Not Available | Not Available | -500 | 2000 |
| 1993-12-01 | SQM | 30 | Not Available | Not Available | -500 | 2000 |
| 1993-12-01 | SQM | 48 | Not Available | Not Available | -500 | 2000 |
| 1993-12-01 | SQM | 21 | Not Available | Not Available | -500 | 2000 |
| 1993-12-01 | SQM | 98 | Not Available | Not Available | -500 | 2000 |
| 1993-12-01 | SQM | 1071 | Not Available | Not Available | -500 | 2000 |
| 1993-12-01 | SQM | 4 | Not Available | Not Available | -500 | 2000 |
| 1993-12-01 | SQM | 8328 | Not Available | Not Available | -500 | 2000 |
| 1993-12-01 | SQM | 8328 | Not Available | Not Available | -500 | 2000 |
| 1993-12-01 | SQM | 291 | Not Available | Not Available | -500 | 2000 |
| 1990-08-01 | Number | 4977 | Not Available | Total Both Sexes | /products/demos | 0 |
| 1993-12-01 | SQM | 306 | Not Available | Not Available | -500 | 2000 |
| 1993-12-01 | SQM | 35 | Not Available | Not Available | -500 | 2000 |
| 1993-12-01 | SQM | 42 | Not Available | Not Available | -500 | 2000 |
|-----|
only showing top 20 rows
|-----|
[20000_dfg] [20000_46454]
-----
| REVISED | Not Available |
| REVISED | Not Available |
| REVISED | Not Available |
| REVISED | Not Available |
| REVISED | Not Available |
| FINAL | Not Available |
| 71 | Not Available |
| 72 | Not Available |
| 73 | Not Available |
| 74 | Not Available |
| 75 | Not Available |
| 76 | Not Available |
| 099999 | Not Available |
| 777 | Not Available |
| 1 | Not Available |
| FINAL | Not Available |
| 2 | Not Available |
| 3 | Not Available |
| 11 | Not Available |
|-----|
only showing top 20 rows
|-----|
[VVWVHGQ_131] [VVWVHGQ_dfg] [VVWVHGQ_585] [VVWVHGQ_cvacc] [VVWVHGQ_13pff] [VVWVHGQ_dfg] [VVWVHGQ_456] [VVWVHGQ_wrcwf]
-----
| 100100 | 1990-04 | New | 807 | 3 | Not Available | Not Available | 2000 |
| 100200 | 1990-05 | New | 5109 | 3 | Not Available | Not Available | 2000 |
| 101100 | 1991-02 | New | 4 | 3 | Not Available | Not Available | 2000 |
| 102100 | 1990-08 | New | 5427 | 3 | Not Available | Not Available | 2000 |
| 200100 | 2014-04 | New | 569 | 3 | Not Available | Not Available | 2000 |
| 100900 | 2007-04 | Altered | 6 | 3 | Not Available | Not Available | 2000 |
| 100900 | 1991-12 | Industrial | 1 | 6 | Labour Force Stat... | Not Available | 6411 |
|-----|

```

```

Terminal
hadoop@qija-Linux-Virtual:
File Edit View Search Terminal Help
|-----|
|-----|
|-----|
only showing top 20 rows
|-----|
[20000_dfg] [20000_46454]
-----
| REVISED | Not Available |
| REVISED | Not Available |
| REVISED | Not Available |
| REVISED | Not Available |
| REVISED | Not Available |
| FINAL | Not Available |
| 71 | Not Available |
| 72 | Not Available |
| 73 | Not Available |
| 74 | Not Available |
| 75 | Not Available |
| 76 | Not Available |
| 099999 | Not Available |
| 777 | Not Available |
| 1 | Not Available |
| FINAL | Not Available |
| 2 | Not Available |
| 3 | Not Available |
| 11 | Not Available |
|-----|
only showing top 20 rows
|-----|
[VVWVHGQ_131] [VVWVHGQ_dfg] [VVWVHGQ_585] [VVWVHGQ_cvacc] [VVWVHGQ_13pff] [VVWVHGQ_dfg] [VVWVHGQ_456] [VVWVHGQ_wrcwf]
-----
| 100100 | 1990-04 | New | 807 | 3 | Not Available | Not Available | 2000 |
| 100200 | 1990-05 | New | 5109 | 3 | Not Available | Not Available | 2000 |
| 101100 | 1991-02 | New | 4 | 3 | Not Available | Not Available | 2000 |
| 102100 | 1990-08 | New | 5427 | 3 | Not Available | Not Available | 2000 |
| 200100 | 2014-04 | New | 569 | 3 | Not Available | Not Available | 2000 |
| 100900 | 2007-04 | Altered | 6 | 3 | Not Available | Not Available | 2000 |
| 100900 | 1991-12 | Industrial | 1 | 6 | Labour Force Stat... | Not Available | 6411 |
|-----|

```

```
Activities Terminal
hadoop@surja-lanovo-G560 ~
File Edit View Search Terminal Help
Tue 18:23
only showing top 28 rows
POPPPIP_3r3g| POPPIP_werw| POPPIP_55|POPPPIP_dfgdg4|POPPPIP_jhk6| POPPIP_gh3|POPPPIP_e3e|POPPPIP_tweT36|
North Cape|BLDM_SV0000001|Building consents...| Not Available|Not Available|Not Available| -500| A1001000|
Rangunu Harbour|BLDM_SV0000001|Building consents...| Not Available|Not Available|Not Available| -500| A1002000|
Tamanunaru|BLDM_SV0000001|Building consents...| Not Available|Not Available|Not Available| -500| A1013000|
Hangawhai|BLDM_SV0000001|Building consents...| Not Available|Not Available|Not Available| -500| A1010000|
Rotorua Central|BLDM_SV0000001|Building consents...| Not Available|Not Available|Not Available| -500| A2138000|
Fairly Springs|BLDM_SV00010001|Building consents...| Not Available|Not Available|Not Available| -500| A3066000|
Inlet Watomo Dis...|BLDM_SV00120004|Building consents...| Not Available|Not Available|Not Available| 17| 204|
Herangi|BLDM_SV00120004|Building consents...| Not Available|Not Available|Not Available| 24| 204|
Hangatiki|BLDM_SV00120004|Building consents...| Not Available|Not Available|Not Available| 11| 204|
Aria|BLDM_SV00120004|Building consents...| Not Available|Not Available|Not Available| 27| 204|
Te Kuiti West|BLDM_SV00120004|Building consents...| Not Available|Not Available|Not Available| 27| 204|
Te Kuiti East|BLDM_SV00120004|Building consents...| Not Available|Not Available|Not Available| 13| 204|
Waipa Valley|BLDM_SV00120004|Building consents...| Not Available|Not Available|Not Available| 17| 204|
Tiroa|BLDM_SV00120004|Building consents...| Not Available|Not Available|Not Available| 39| 204|
Haroti|BLDM_SV00120004|Building consents...| Not Available|Not Available|Not Available| 31| 204|
Ohakuri|BLDM_SV00120004|Building consents...| Not Available|Not Available|Not Available| 2| 204|
Western Bay of |BLDM_SV00120004|Building consents...| Not Available|Not Available|Not Available| -500| A0000000|
Lake Taupo Bay|BLDM_SV00120004|Building consents...| Not Available|Not Available|Not Available| 15| 204|
Napua|BLDM_SV00120004|Building consents...| Not Available|Not Available|Not Available| 0| 204|
Inland water Lake...|BLDM_SV00120004|Building consents...| Not Available|Not Available|Not Available| 15| 204|
only showing top 30 rows
BNVVBH_3453|BNVVBH_g3g|BNVVBH_dfgd| BNVVBH_yfa5| BNVVBH_er|BNVVBH_56urg| BNVVBH_tryut|BNVVBH_sdfgs| BNVVBH_
456|BNVVBH_3em|
Far north District| 0| F|Building consents...|All construction| 1| Not Available|Not Available|Not Avail
```

```
Activities Terminal
hadoop@surja-lanovo-G560 ~
File Edit View Search Terminal Help
Tue 18:23
only showing top 28 rows
BNVVBH_3453|BNVVBH_g3g|BNVVBH_dfgd| BNVVBH_yfa5| BNVVBH_er|BNVVBH_56urg| BNVVBH_tryut|BNVVBH_sdfgs| BNVVBH_
456|BNVVBH_3em|
Far north District| 0| F|Building consents...|All construction| 1| Not Available|Not Available|Not Avail
ible| 9| 204|
Far north District| 0| F|Building consents...|All construction| 2| Not Available|Not Available|Not Avail
ible| 198| 204|
Far north District| 0| F|Building Consents...|All construction| 13| Not Available|Not Available|Not Avail
ible| 39| 204|
Walpara District| 0| F|Building Consents...|All construction| 8| Not Available|Not Available|Not Avail
ible| 12| 204|
Rotorua District| 0| F|Building Consents...|All construction| 15| Not Available|Not Available|Not Avail
ible| 0| 204|
Rotorua District| 0| F|Building Consents...| Dwelling units| 7| Not Available|Not Available|Not Avail
ible| 0| 204|
Watomo District| 0| F|Building Consents...| Factories| 2|Household Labour ...|Not Available|Not Avail
ible| A237200| 2|Household Labour ...|Not Available|Not Avail
ible| A237000| 2|Household Labour ...|Not Available|Not Avail
ible| A237000| 2|Household Labour ...|Not Available|Not Avail
ible| A237000| 2|Household Labour ...|Not Available|Not Avail
ible| A238000| 2|Household Labour ...|Not Available|Not Avail
ible| A238200| 2|Household Labour ...|Not Available|Not Avail
ible| A238400| 2|Household Labour ...|Not Available|Not Avail
ible| A238500| 2|Household Labour ...|Not Available|Not Avail
ible| A238500| 2|Household Labour ...|Not Available|Not Avail
```

```
Activities Terminal
hadoop@surja-lanovo-G560 ~
File Edit View Search Terminal Help
Tue 18:23
only showing top 28 rows
DERDOF_3| DERDOF_dgd| DERDOF_fgh| DERDOF_kthf| DERDOF_4645|DERDOF_eyey|
0|Not Available|Not Available|Not Available|Not Available| 35|
0|Not Available|Not Available|Not Available|Not Available| 33|
0|Not Available|Not Available|Not Available|Not Available| 33|
0|Not Available|Not Available|Not Available|Not Available| 21|
0|Not Available|Not Available|Not Available|Not Available| 21|
0|Not Available|Not Available|Not Available|Not Available| 6|
0|Not Available|Not Available|Not Available|Not Available| 2012|
0|Not Available|Not Available|Not Available|Not Available| 2012|
0|Not Available|Not Available|Not Available|Not Available| 2012|
0|Not Available|Not Available|Not Available|Not Available| 2012|
0|Not Available|Not Available|Not Available|Not Available| 2012|
0|Not Available|Not Available|Not Available|Not Available| 2012|
0|Not Available|Not Available|Not Available|Not Available| 2012|
0|Not Available|Not Available|Not Available|Not Available| 2012|
0|Not Available|Not Available|Not Available|Not Available| 2012|
0|Not Available|Not Available|Not Available|Not Available| 2012|
0|Not Available|Not Available|Not Available|Not Available| 2012|
0|Not Available|Not Available|Not Available|Not Available| 52|
0|Not Available|Not Available|Not Available|Not Available| 2012|
0|Not Available|Not Available|Not Available|Not Available| 2012|
0|Not Available|Not Available|Not Available|Not Available| 2012|
only showing top 28 rows
VRYVRC_dg|VRYVRC_dyutydfag| VRYVRC_4564|VRYVRC_egvev|
HLFQ_S1A15| 1986-03|Not Available| 45|
HLFQ_S1A15| 1986-06|Not Available| 51|
HLFQ_S1A15| 1988-09|Not Available| 43|
HLFQ_S1A15| 2011-03|Not Available| 0|
HLFQ_S1A35| 2001-03|Not Available| 38|
HLFQ_S2A3Q2| 1986-12|Not Available| 10|
FINAL| 3098765|Not Available| 10|
```



```

Activities Terminal * Tue 10:24 *
hadop@surja-Lenovo-G56E -
File Edit View Search Terminal Help
-----
NNNNNN_rtyr| NNNNNN_bdb| NNNNNN_ee| NNNNNN_rhf|
Household Labour ... |Not Available|Not Available|Not Available|
Household Labour ... |Not Available|Not Available|Not Available|
Household Labour ... |Not Available|Not Available|Not Available|
Household Labour ... |Not Available|Not Available|Not Available|
Household Labour ... |Not Available|Not Available|Not Available|
Household Labour ... |Not Available|Not Available|Not Available|
Percent|Not Available|Not Available|Not Available|
Percent|Not Available|Not Available|Not Available|
Percent|Not Available|Not Available|Not Available|
Percent|Not Available|Not Available|Not Available|
Percent|Not Available|Not Available|Not Available|
Percent|Not Available|Not Available|Not Available|
Percent|Not Available|Not Available|Not Available|
Percent|Not Available|Not Available|Not Available|
Percent|Not Available|Not Available|Not Available|
Percent|Not Available|Not Available|Not Available|
Household Labour ... |Not Available|Not Available|Not Available|
Percent|Not Available|Not Available|Not Available|
Percent|Not Available|Not Available|Not Available|
Percent|Not Available|Not Available|Not Available|
-----
only showing top 20 rows
-----
[EEREHWRED_dsfqsdq|EEREHWRED_tytut|EEREHWRED_s0vs| EEREHWRED_f|EEREHWRED_s0ysh|EEREHWRED_dfgs|EEREHWRED_egb|EEREHWRED_3453| EEREHWRED_ett3w
T]
-----
0| 1973| Not Available|Not Available| Not Available| Not Available| 2012| 0| /product/product
0| 2591| Not Available|Not Available| Not Available| Not Available| 2011| 0| /den
0| 1525| Not Available|Not Available| Not Available| Not Available| 2012| 0| /download/downloa...

```

```

Activities Terminal * Tue 10:24 *
hadop@surja-Lenovo-G56E -
File Edit View Search Terminal Help
-----
New| HLFQ_52H3Q8|Not Available|Not Available|Not Available| heavydog| 0|
New| HLFQ_52H3Q8|Not Available|Not Available|Not Available| heavydog| 0|
-----
only showing top 20 rows
-----
TRCHTT_w|TRCHTT_bdd|TRCHTT_audre| TRCHTT_gdfgs|TRCHTT_dfta|
New Zealand| 0| Number|Not Available| 18|
New Zealand| 0| Number|Not Available| 19|
New Zealand| 0| Number|Not Available| 19|
New Zealand| 0| Number|Not Available| 7|
New Zealand| 1| Number|Not Available| 7|
New Zealand| 19| Number|Not Available| 18|
Far North District|Floor area| 1996.12|Not Available| 7|
Far North District|Floor area| 2000.03|Not Available| 2|
Far North District|Floor area| 2000.80|Not Available| 6|
Far North District|Floor area| 2000.09|Not Available| 2|
Far North District|Floor area| 2000.12|Not Available| 6|
Far North District|Floor area| 2001.03|Not Available| 2|
Far North District|Floor area| 2001.06|Not Available| 6|
Far North District|Floor area| 2001.09|Not Available| 2|
Far North District|Floor area| 2001.12|Not Available| 6|
Far North District|Floor area| 2002.03|Not Available| 2|
New Zealand| 19| Number|Not Available| 21|
Far North District|Floor area| 2002.06|Not Available| 6|
Far North District|Floor area| 2002.09|Not Available| 6|
Far North District|Floor area| 2002.12|Not Available| 6|
-----
only showing top 20 rows
-----
[UITT3QG_4354| UITT3QG_rfgbrs| UITT3QG_rst|UITT3QG_gkshst|
|Not Available| Not Available|Not Available| 1743|
|Not Available| Not Available|Not Available| 75756|
|Not Available| Not Available|Not Available| 55395|
|Not Available| Not Available|Not Available| 35451|

```

```

Activities Terminal * Tue 10:24 *
hadop@surja-Lenovo-G56E -
File Edit View Search Terminal Help
-----
NNNNNN_rtyr| NNNNNN_bdb| NNNNNN_ee| NNNNNN_rhf|
Household Labour ... |Not Available|Not Available|Not Available|
Household Labour ... |Not Available|Not Available|Not Available|
Household Labour ... |Not Available|Not Available|Not Available|
Household Labour ... |Not Available|Not Available|Not Available|
Household Labour ... |Not Available|Not Available|Not Available|
Household Labour ... |Not Available|Not Available|Not Available|
Percent|Not Available|Not Available|Not Available|
Percent|Not Available|Not Available|Not Available|
Percent|Not Available|Not Available|Not Available|
Percent|Not Available|Not Available|Not Available|
Percent|Not Available|Not Available|Not Available|
Percent|Not Available|Not Available|Not Available|
Percent|Not Available|Not Available|Not Available|
Percent|Not Available|Not Available|Not Available|
Percent|Not Available|Not Available|Not Available|
Household Labour ... |Not Available|Not Available|Not Available|
Percent|Not Available|Not Available|Not Available|
Percent|Not Available|Not Available|Not Available|
Percent|Not Available|Not Available|Not Available|
-----
only showing top 20 rows
-----
[EEREHWRED_dsfqsdq|EEREHWRED_tytut|EEREHWRED_s0vs| EEREHWRED_f|EEREHWRED_s0ysh|EEREHWRED_dfgs|EEREHWRED_egb|EEREHWRED_3453| EEREHWRED_ett3w
T]
-----
0| 1973| Not Available|Not Available| Not Available| Not Available| 2012| 0| /product/product
0| 2591| Not Available|Not Available| Not Available| Not Available| 2011| 0| /den
0| 1525| Not Available|Not Available| Not Available| Not Available| 2012| 0| /download/downloa...

```

CONCLUSION

The system is flexible and the changes if any can be made without much difficulty .Every step has been taken to make the working of the project comfortable for the users. Data Analytics Done by Real Time Data using Hadoop and Spark is a Business Analytics Application. Its main goal is to migrate and transfer the data from client and Application server to HIVE with the help of Spark. This Application is generally handled by Data Engineer. Develop the data mindset and analytical skills to make informed business decisions here the main decision which I made is deals with data which can be migrated and transfer using Avro format and finally comes in parquet format which is default format of spark is dump in hive. Generally we have Client

Server, Application Server, Spark and Hive. My Project Deals with these software's only. In Client Server data is stored in form of Key Value pair and Maximum deals with MySQL. In Application Server is a Key Data. Spark is used for the speed access and manipulate of the data with the help of .jar files .Spark make use of HBASE to convert Avro format to data frames .These Data Frames are dump into HIVE in tabular format (parquet). Benefit is Client can monitor the transfer through log files. Finally internal data serialization and data migration is done.

BOOKS:

“Hadoop the definition guide” by Tom White

“Learning Spark Lighting fast data analysis” by Holden Karau

“Programming Hive “by Edward Capriole

“HBase the Definitive Guide “By Lars George

WEBSITES:

www.digitalocean.com

www.phoenixnap.com

<https://hbase.apache.org/>

<https://hadoop.apache.org/>

<https://spark.apache.org/docs/latest/>

<https://hive.apache.org/>

Real-time data processing scheme using big data analytics in internet of things based smart transportation environment

M Babar, F Arif - Journal of Ambient Intelligence and Humanized ..., 2019 - Springer

A survey on real-time big data analytics: applications and tools

B Yadrnjiaghdam, N Pool... - ... on computational science ..., 2016 - ieeexplore.ieee.org

