Name	Pranay Singhvi
UID No.	202130026

Experiment 7

HONOR PLEDGE I hereby declare that the documentation, code and output attached with this lab exporiment has been completed by me in accordance with highest standards of honesty. I confirm that I have not plagiarized og used unauthorized natorial or given or received illegitimate help for completing this experiment. I will uphold equity and honesty in the evalutation on my work, and if found quilty of plagiarism or dishonesty, will bear the consequences as outlined in the integrity section of the lab substice. I am doing so in order to maintain

PROBLEM STATEMENT

Title: Big data analysis using Hive/ Use of Graph database

- 1. Import the batch-specific data and store it as a Hive Table
- 2. Perform a Hive guery on your uploaded dataset
- 3. Pull data into a spark dataframe and repeat the query using a spark dataframe
- 4. Link neo4j to your underlying datastore and run a graph query

THEORY

1. Hadoop Overview:

Hadoop is an open-source framework designed for distributed storage and processing of large data sets across clusters of commodity hardware. It provides a distributed file system (HDFS) for storage and a framework (MapReduce) for processing and analyzing large datasets. Installation:

a community built around this code of honous

Guide Followed: Hadoop-on-Ubuntu

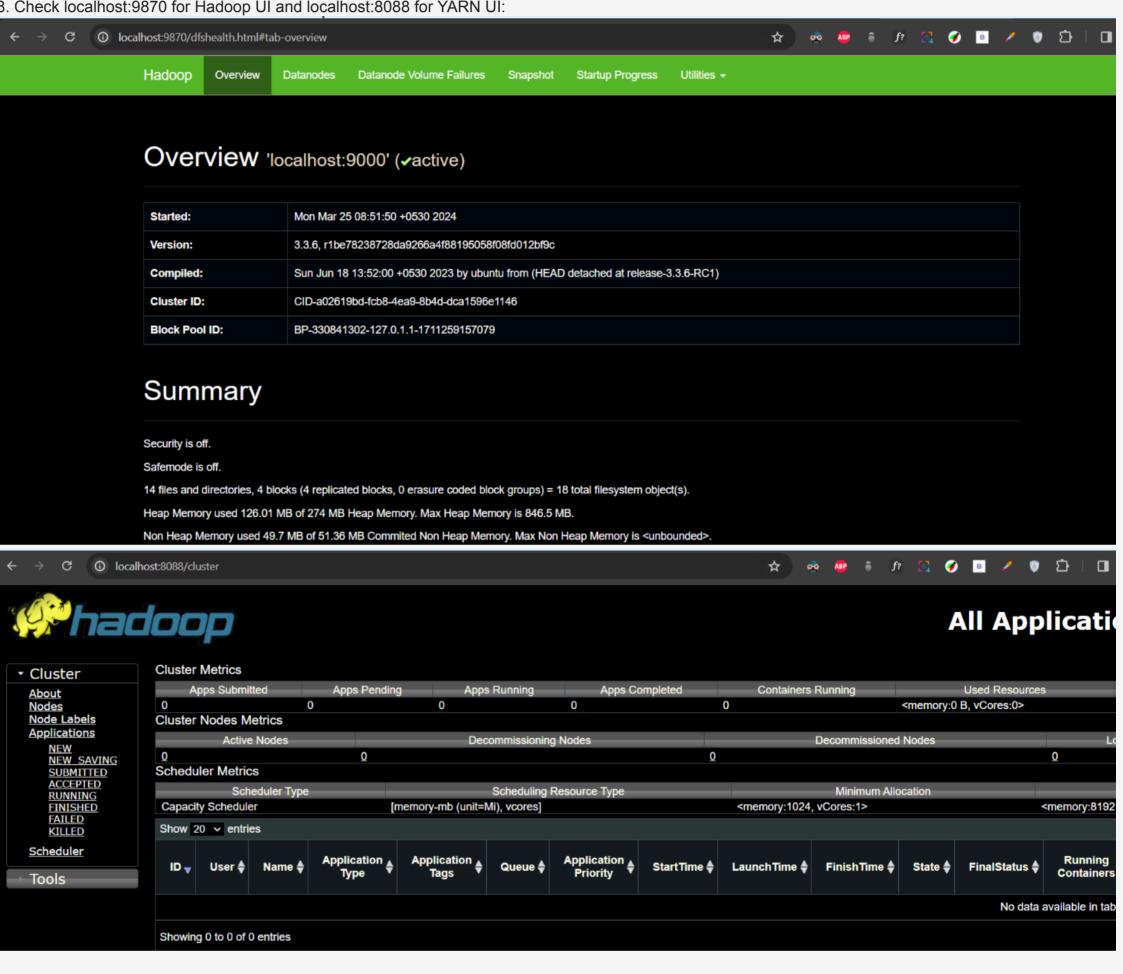
- 1. Create New User for Hadoop: sudo adduser hdoop & su hdoop
- 2. Set Up SSH Kevs: ssh-kevgen -t rsa -P " -f ~/.ssh/id rsa
- 3. Download the latest stable release of Hadoop from the Apache Hadoop Website.
- 4. Configure Hadoop Environment: Set JAVA HOME in /usr/local/hadoop/etc/hadoop/hadoop-env.sh
- 5. Update Hadoop's XML configuration files with default config:

6. Initialize the Hadoop filesystem namespace: /usr/local/hadoop/bin/hdfs namenode -format

hdoop@DESKTOP-5UL8JKF:~\$ start-dfs.sh Starting namenodes on [localhost] Starting datanodes Starting secondary namenodes [DESKTOP-5UL8JKF] hdoop@DESKTOP-5UL8JKF:~\$ start-yarn.sh Starting resourcemanager Starting nodemanagers hdoop@DESKTOP-5UL8JKF:~\$

7. Start Hadoop, Launch HDFS and YARN services: start-all.sh, verify using: jps

hdoop@DESKTOP-5UL8JKF:~\$ jps
64678 Jps
935 NameNode
1720 NodeManager
1242 SecondaryNameNode
1051 DataNode
1598 ResourceManager
hdoop@DESKTOP-5UL8JKF:~\$
8. Check localhost:9870 for Hadoop UI and localhost:8088 for YARN UI:



2. Hive Overview:

enron

hive>

temp_employee

Hive is a data warehouse infrastructure built on top of Hadoop for providing data summarization, query, and analysis. It provides a SQL-like interface (HiveQL) to query and analyze data stored in Hadoop's distributed file system (HDFS).

Installation:

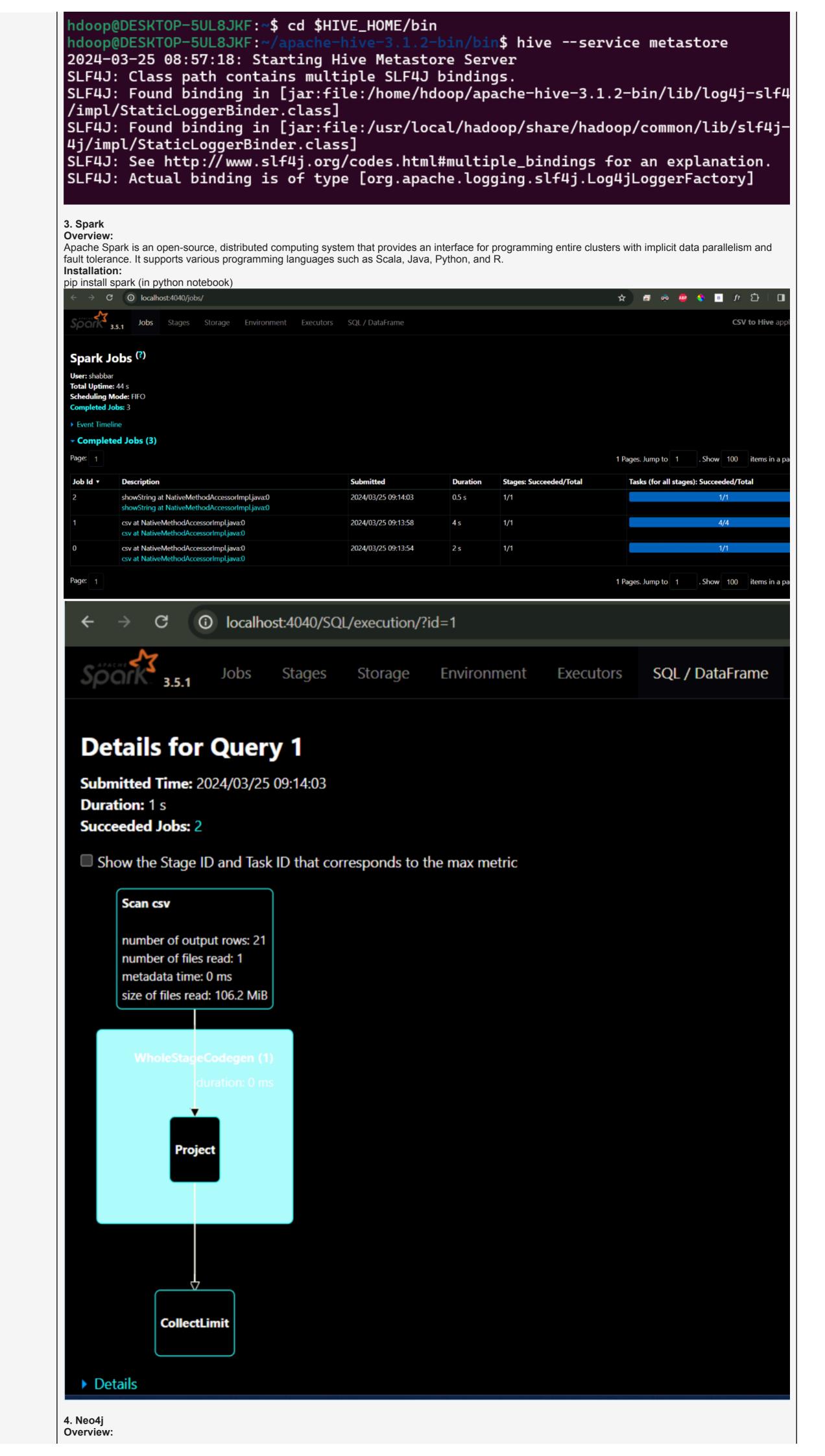
Guide Followed: Hive-on-Ubuntu

- 1. Download the latest stable release of Hive from the Apache Hive website.
- 2. Extract the downloaded archive to a directory on your WSL Ubuntu system.

Time taken: 1.668 seconds, Fetched: 2 row(s)

- 3. Set up environment variables in your `.bashrc` file to specify the Hive home directory and add Hive's bin directory to your PATH.
- 4. Configure Hive's XML files ('hive-site.xml', 'hive-default.xml', etc.) according to your Hadoop and metastore setup.
- 5. Create Hive Directories in HDFS, Create two separate directories to store data in the HDFS layer: tmp & warehouse 6. Initiate Derby Database:
- \$HIVE_HOME/bin/schematool -dbType derby -initSchema 7. Start Hive services and the metastore using the provided scripts ('hive --service metastore', etc.).

```
hdoop@DESKTOP-5UL8JKF:~/apache-hive-3.1.2-bin/bin$ hive
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/hdoop/apache-hive-3.1.2-bin/lib/log4j-slf4j-impl-2.10.
/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/local/hadoop/share/hadoop/common/lib/slf4j-reload4j-1.7
4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.apache.logging.slf4j.Log4jLoggerFactory]
Hive Session ID = bc63537d-dd02-4292-806a-2bafc4bcee4f
Logging initialized using configuration in jar:file:/home/hdoop/apache-hive-3.1.2-bin/lib/hive
r!/hive-log4j2.properties Async: true
Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider (
 execution engine (i.e. spark, tez) or using Hive 1.X releases.
Hive Session ID = b7729195-232a-434e-ae7e-293192d9374d
hive> show tables;
OK
```



```
Neo4j is a graph database management system that provides an efficient and expressive way to store, manage, and query highly connected data. It is
optimized for handling graph data and supports powerful querying capabilities using the Cypher query language.
Guide Followed: Neo4j-on-Ubuntu
1. Download the latest stable release of Neo4j from the Neo4j Website
2. Follow the installation instructions provided for Linux distributions.
3. After installation, start the Neo4j service using the provided startup scripts: sudo systematl enable/start/edit/status/stop neo4j.
4. Configure username and password through cypher-shell: :server change-password
5. Access the Neo4j browser interface using a web browser and connect to the Neo4j database.
hdoop@DESKTOP-5UL8JKF:~$ sudo systemctl restart neo4j
hdoop@DESKTOP-5UL8JKF:~$ cypher-shell -a 'neo4j://127.0.0.1:7687'
username: neo4j
password:
Password change required
new password:
confirm password:
hdoop@DESKTOP-5UL8JKF:~$ cypher-shell -a 'neo4j://127.0.0.1:7687'
username: neo4j
password:
Password change required
new password:
confirm password:
Connected to Neo4j using Bolt protocol version 5.4 at neo4j://127.0.0.1:7
Type :help for a list of available commands or :exit to exit the shell.
Note that Cypher queries must end with a semicolon.
neo4j@neo4j>
hdoop@DESKTOP-5UL8JKF:~$ cypher-shell -a 'neo4j://127.0.0.1:7687'
username: neo4j
password:
Connected to Neo4j using Bolt protocol version 5.4 at neo4j://127.0.0.1:7687 as user neo4j.
Type :help for a list of available commands or :exit to exit the shell.
Note that Cypher queries must end with a semicolon.
neo4j@neo4j> CREATE (:Shark {name: 'Great White'});
0 rows
ready to start consuming query after 542 ms, results consumed after another 0 ms
Added 1 nodes, Set 1 properties, Added 1 labels
neo4j@neo4j> CREATE
               (:Shark {name: 'Hammerhead'})-[:FRIEND]→
               (:Shark {name: 'Sammy'})-[:FRIEND]→
               (:Shark {name: 'Megalodon'});
0 rows
ready to start consuming query after 250 ms, results consumed after another 0 ms
Added 3 nodes, Created 2 relationships, Set 3 properties, Added 3 labels
neo4j@neo4j> MATCH (a:Shark),(b:Shark)
              WHERE a.name = 'Sammy' AND b.name = 'Megalodon'
              CREATE (a)-[r:ORDER { name: 'Lamniformes' }]\rightarrow(b)
              RETURN type(r), r.name;
  type(r) | r.name
   "ORDER"
             "Lamniformes"
```

1. Importing Libraries & Dataset

```
import re
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from neo4j import GraphDatabase
from pyspark.sql import SparkSession
# read csv file
df = pd.read_csv('../Datasets/enron.csv', index_col=0, low_memory=False)
df.head()
```

t[]:	Message-ID Date F		From	То		ubject X- From		Х-То		X- bcc	X-F	
ut[]: - (0	<18782981.1075855378110.JavaMail.evans@thyme>	2001- 05-14 23:39:00	frozenset({'phillip.allen@enron.com'})	frozenset({'tim.belden@enron.com'}))	NaN	Phillip K Allen	Tim Belden <tim Belden/Enron@EnronXGate></tim 	NaN	NaN	\Phillip_Allen_Jan2002_1\ Phillip K.\
	1	<15464986.1075855378456.JavaMail.evans@thyme>	2001- 05-04 20:51:00	frozenset({'phillip.allen@enron.com'})	frozenset({'john.lavorato@enron.com'}))	Re:	Phillip K Allen	John J Lavorato <john j<br="">Lavorato/ENRON@enronXg</john>	NaN	NaN	\Phillip_Allen_Jan2002_1\ Phillip K.\
	2	<24216240.1075855687451.JavaMail.evans@thyme>	2000- 10-18 10:00:00	frozenset({'phillip.allen@enron.com'})	frozenset({'leah.arsdall@enron.com'})) R		Phillip K Allen	Leah Van Arsdall	NaN	NaN	\Phillip_Allen_Dec2000\t Folders\'sen
	3	<13505866.1075863688222.JavaMail.evans@thyme>	2000- 10-23 13:13:00	frozenset({'phillip.allen@enron.com'})	frozenset({'randall.gay@enron.com'}))	NaN	Phillip K Allen	Randall L Gay	NaN	NaN	\Phillip_Allen_Dec2000\f Folders\'sen
	4	<30922949.1075863688243.JavaMail.evans@thyme>	2000- 08-31 12:07:00	frozenset({'phillip.allen@enron.com'})	frozenset({'greg.piper@enron.com'}))	Re: Hello	Phillip K Allen	Greg Piper	NaN	NaN	\Phillip_Allen_Dec2000\f Folders\'sen

Out

2. Preprocessing Data

```
In []: def sanitize_column_name(name):
    return re.sub(r'[^A-Za-za-9_]', '_', name)
    new_columns = [sanitize_column_name(col) for col in df.columns]
    df.columns = new_columns
    df.co_csv("../Datasets/cleaned_enron.csv", index=False)

In []: df_new = pd.read_csv('../Datasets/cleaned_enron.csv', index_col=0, low_memory=False)
    df_subset = pd.fi.loc[:50000]
    df_subset = pd.read_csv('../Datasets/enron_chunk.csv", index_False)

In []: df_subset = pd.read_csv('../Datasets/enron_chunk.csv', index_col=0, low_memory=False)
    print(df_subset.shape[0])
    df_subset = pd.read_csv('../Datasets/enron_chunk.csv', index_col=0, low_memory=False)
    df_subset = pd.read_csv('../Datasets
```

:		Date	From	То	Subj	ect	X_From	X_To	X_cc	X_bcc	
	Message_ID										
	<18782981.1075855378110.JavaMail.evans@thyme>	2001- 05-14 23:39:00	frozenset({'phillip.allen@enron.com'})	frozenset({'tim.belden@enron.com'})	Ν	NaN	Phillip K Allen	Tim Belden <tim Belden/Enron@EnronXGate></tim 	NaN	NaN	\Phillip_Allen_Jan20 Ph
	<15464986.1075855378456.JavaMail.evans@thyme>	2001- 05-04 20:51:00	frozenset({'phillip.allen@enron.com'})	frozenset({'john.lavorato@enron.com'})		Re:	Phillip K Allen	John J Lavorato < John J Lavorato/ENRON@enronXg	NaN	NaN	\Phillip_Allen_Jan20 Ph
	<24216240.1075855687451.JavaMail.evans@thyme>	2000- 10-18 10:00:00	frozenset({'phillip.allen@enron.com'})	frozenset({'leah.arsdall@enron.com'})	Re: t	test	Phillip K Allen	Leah Van Arsdall	NaN	NaN	\Phillip_Allen_Dec Folder
	<13505866.1075863688222.JavaMail.evans@thyme>	2000- 10-23 13:13:00	frozenset({'phillip.allen@enron.com'})	frozenset({'randall.gay@enron.com'})	Ν	NaN	Phillip K Allen	Randall L Gay	NaN	NaN	\Phillip_Allen_Dec Folder
	<30922949.1075863688243.JavaMail.evans@thyme>	2000- 08-31 12:07:00	frozenset({'phillip.allen@enron.com'})	frozenset({'greg.piper@enron.com'})		Re: ello	Phillip K Allen	Greg Piper	NaN	NaN	\Phillip_Allen_Dec Folder

5 rows × 50 columns

3. Loading Dataset to a Spark Dataframe

```
In [ ]: # Create SparkSession
        spark = SparkSession.builder \
            .appName("CSV to Hive") \
            .config("hive.metastore.uris", "thrift://localhost:9083") \
            .enableHiveSupport() \
            .get0rCreate()
        # Read CSV file into DataFrame
        csv_file_path = "./eron2.csv"
        df1 = spark.read.csv(csv_file_path, header=True, inferSchema=True)
        # Define a function to sanitize column names
        # def sanitize_column_name(name):
              return re.sub(r'[^A-Za-z0-9_]', '_', name)
        # new_columns = [sanitize_column_name(col) for col in df1.columns]
        # df1 = df1.toDF(*new_columns)
        df1.show()
       your 131072x1 screen size is bogus. expect trouble
       Warning: Ignoring non-Spark config property: hive.metastore.uris
       24/03/24 11:55:07 WARN Utils: Your hostname, DESKTOP-5UL8JKF resolves to a loopback address: 127.0.1.1; using 172.27.122.177 instead (on interface eth0)
       24/03/24 11:55:07 WARN Utils: Set SPARK_LOCAL_IP if you need to bind to another address
       Setting default log level to "WARN".
      To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
      24/03/24 11:55:10 WARN NativeCodeLoader: Unable to load native—hadoop library for your platform... using builtin—java classes where applicable
      24/03/24 11:55:30 WARN SparkStringUtils: Truncated the string representation of a plan since it was too large. This behavior can be adjusted by setting 'spark.sql.debug.maxToStringField
      s'.
```

								++
+-		+	+	+			-++	V colV book
er	Message_ID X_Origin	Date X_FileName	From content	user Cat	_1_level_1	X_From Cat_1_level_2		
	_weight Cat_6_leve		ight Cat_7_level_1 C	at_7_level_2 Cat_7_we	ight Cat_8_level	_1 Cat_8_level_2 Cat_8	_weight Cat_9_level_	_5_level_1 Cat_5_level_2 1 Cat_9_level_2 Cat_9_weight Cat
_10_le	evel_1 Cat_10_level_2 Cat 	_10_weight Cat_11_level_1 	Cat_11_level_2 Cat_1 +	1_weight Cat_12_level 	_1 Cat_12_level_ 	2 Cat_12_weight labele 	d +	
	+					+	+	++
+-		+		+		+	++ -++	
<1878 a	•	14 23:39:00 frozenset({'ph en (Non-Privi Here is o	•	im.b allen-p	NULL Phillip NULL	K Allen Tim Belden <ti NULL </ti 	m B NULL	NULL NULL \Phillip_Allen_J NULL
NULL NULL	NULL NULL	NULL NULL NULL	NULL NULL	NULL NULL NULL	NULL NULL	NULL NULL NULL	NULL NULL	NULL NULL NULL
ULL <1546	•	LL NULL 04	NULL NULL ill frozenset({'j		NULL Re: Phillip	NULL False K Allen John J Lavorat	0 <	NULL NULL \Phillip_Allen_J
a NULL	Allen-P pall NULL	en (Non-Privi Traveling NULL	to have NULL	allen-p NULL	NULL NULL	NULL NULL	NULL NULL	NULL NULL NULL
NULL ULL	NULL NULL NU	NULL NULL LL NULL	NULL NULL NULL	NULL NULL NULL	NULL NULL	NULL NULL NULL False	NULL	NULL NULL N
	.6240.1075855 2000-10- Allen-P	18 10:00:00 frozenset({'ph: pallen.nsf test succ	•	eah R allen-p	e: test Phillip NULL	K Allen Leah Van A NULL	rsdall NULL	<pre>NULL NULL \Phillip_Allen_D</pre>
NULL NULL	NULL NULL	NULL NULL NULL	NULL	NULL NULL NULL	NULL NULL	NULL NULL NULL	NULL NULL	NULL . NULL NULL N
ULL	NULL NU	LL NULL 23 13:13:00 frozenset({'ph	NULL NULL	NULL	NULL NULL Phillip	NULL False	•	NULL NULL \Phillip_Allen_D
e NULL	Allen-P NULL	pallen.nsf Randy, Car NULL	•	alien-p NULL	NULL NULL	NULL NULL	NULL NULL	NULL NULL
NULL ULL	NULL	NULL NULL LL NULL	·	NULL NULL	NULL NULL	NULL NULL NULL False	•	NULL NULL N
	•	31 12:07:00 frozenset({'ph pallen.nsf Let's sho	ill frozenset({'g		: Hello Phillip		Piper NULL	<pre>NULL NULL \Phillip_Allen_D</pre>
NULL	NULL NULL	NULL NULL	NULL	NULL NULL	NULL NULL	NULL NULL NULL	NULL	NULL NULL NULL
ULL	NULL NU	LL NULL 31 11:17:00 frozenset({'ph:	NULL NULL	NULL	NULL : Hello Phillip	NULL False	Piper	NULL NULL \Phillip_Allen_D
e NULL	Allen-P NULL	pallen.nsf Greg, How		allen-p NULL	NULL NULL	NULL NULL	NULL NULL	NULL NULL
NULL ULL	NULL	NULL NULL LL NULL	'	NULL NULL	NULL NULL	NULL NULL NULL False	NULL	NULL NULL N
	•	22 14:44:00 frozenset({'ph. pallen.nsf Please cc	ill frozenset({'d		•	K Allen david.l.johnso NULL	n@e NULL	<pre>NULL NULL \Phillip_Allen_D</pre>
NULL	NULL NULL	NULL NULL	NULL	NULL NULL	NULL NULL	NULL NULL NULL	NULL	NULL NULL NULL
ULL	NULL NU	LL NULL 14 13:59:00 frozenset({'ph:	NULL NULL	NULL	NULL	NULL False	·	NULL NULL \Phillip_Allen_D
e NULL	Allen-P NULL	pallen.nsf any mornin		allen-p NULL	NULL NULL	NULL NULL	NULL NULL	NULL NULL
NULL ULL	NULL	NULL NULL LL NULL	•	NULL NULL	NULL NULL	NULL NULL NULL False	•	NULL NULL N
•	•	17 09:26:00 frozenset({'ph: pallen.nsf 1. login:	ill frozenset({'m	ark Re: High Spee		·	Scott NULL	NULL NULL \Phillip_Allen_D NULL
NULL	NULL NULL	NULL NULL NULL	NULL	NULL NULL NULL	NULL NULL	NULL NULL NULL	NULL	NULL NULL NULL
ULL	NULL NU	LL NULL 16 13:44:00 frozenset({'ph	NULL NULL	NULL	NULL	NULL False	·	NULL NULL \Phillip_Allen_D
e	Allen-P	pallen.nsf "	Buck""	<buck.buc > As di</buck.buc >	scussed d In	a Parallon 75 I a	m developing I	
c. > 8	3725 Pan NM 87113 > NULL NUL	505-7 allen-p		ULL NULL	NULL	NULL NULL NULL NULL	NULL	NULL NULL NUL
	•	16 13:42:00 frozenset({'ph	'	uck Re: FW: fixed	'		uckner Buck"" <buck NULL </buck 	.buc NULL NU NULL
NULL	NULL NULL	NULL NULL	NULL	NULL NULL	NULL NULL	NULL NULL NULL	NULL	NULL NULL NULL
ULL	NULL NU	LL NULL 13 13:45:00 frozenset({'ph	NULL NULL	NULL	NULL	NULL NULL K Allen stagecoachmama	·	NULL NULL \Phillip_Allen_D
e NULL	•	pallen.nsf Lucy, Hero	•	allen-p NULL			NULL NULL	
NULL ULL	NULL	NULL NULL LL NULL	•	NULL NULL	NULL NULL	NULL NULL NULL False	•	NULL NULL N
	·	09 14:16:00 frozenset({'ph. pallen.nsf	ill frozenset({'k		•		Holst NULL	<pre>NULL NULL \Phillip_Allen_D</pre>
NULL	NULL NULL	NULL NULL NULL	NULL	NULL NULL	NULL NULL	NULL NULL NULL	NULL	NULL NULL NULL
ULL	NULL NU	LL NULL 09 14:00:00 frozenset({'ph:	NULL NULL	NULL	NULL	NULL False	Holst	NULL NULL \Phillip_Allen_D
e NULL	Allen-P NULL	pallen.nsf	-	allen-p NULL	NULL NULL	NULL NULL	NULL NULL	NULL NULL NULL
NULL ULL	NULL	NULL NULL LL NULL	·	NULL NULL	NULL NULL	NULL NULL NULL False	•	NULL NULL N
•	-	05 13:26:00 frozenset({'ph pallen.nsf Dave, Hero	ill frozenset({'d		NULL Phillip NULL		lainey NULL	NULL NULL \Phillip_Allen_D NULL
NULL	NULL NULL	NULL NULL	NULL	NULL	NULL NULL	NULL NULL NULL	NULL	NULL NULL NULL
NULL ULL I<1115	NULL NU	NULL NULL LL NULL 05 12:55:00 frozenset({'ph:	NULL NULL	NULL	NULL	NULL False	NULL Harris	NULL NULL \Phillip_Allen_D
e NULL	Allen-P NULL				NULL NULL	NULL NULL	NULL NULL	NULL NULL NULL
NULL	NULL	NULL NULL	•	NULL NULL	NULL NULL	NULL NULL NULL False	•	NULL NULL N
٠.	-	LL NULL 04 16:23:00 frozenset({'ph:	ill frozenset({'i	na.r Var, Reportin	g an Phillip	K Allen Ina	Rangel NULL	NULL NULL \Phillip_Allen_D
e NULL	NULL	pallen.nsf	NULL	allen-p NULL NULL NULL	NULL NULL	NULL NULL	NULL	NULL NULL NULL
NULL ULL	•	NULL NULL LL NULL	NULL NULL	NULL	NULL NULL	NULL NULL NULL False	•	NULL NULL N
a	Allen-P pall	04 18:26:00 frozenset({'ph en (Non-Privi Tim, mike	grigsby	allen-p	NULL	K Allen Tim Heizenrade NULL	NULL	NULL NULL \Phillip_Allen_J NULL
NULL	NULL NULL	NULL NULL NULL	•	NULL NULL NULL	NULL NULL	NULL NULL NULL	NULL NULL	NULL NULL NULL NULL N
•	9088.1075855 2000-10-	LL NULL 03 16:30:00 frozenset({'ph:		alle W		• •	•	NULL NULL \Phillip_Allen_D
	please let me kn v	ery short windo but it	would be 000 to	secure the Georg	e W. Richar	Creekside Builders L	LC - winmail.dat"	t by individua just as the d allen-p NUL
L NULL	•	NULL NULL NULL NULL	NULL NU	LL NULL	NULL	NULL . NUL	NULL NULL L NULL	NULL NULL
e	Allen-P	03 16:15:00 frozenset({'ph: pallen.nsf		allen-p	NULL	NULL	Rangel NULL	NULL NULL \Phillip_Allen_D NULL
NULL	NULL NULL	NULL NULL NULL	•	NULL NULL NULL	NULL NULL	NULL NULL NULL	NULL NULL	NULL NULL NULL NULL N
•	·+	•	·	+	·	•	·	++
+-	· +	· +	· +	· +	· +	· +	+	++
	·	+	·		·	·		

4. Using Spark to load Data into Hive Table

StructType([StructField('Message_ID', StringType(), True), StructField('Date', StringType(), True), StructField('From', StringType(), True), StructField('Y_From', StringType(), True), StructField('Y_To', StringType(), True), StructField('Y_From', StringType(), True), StructField('Cat_1_level_1', StringType(), True), StructField('Cat_2_level_1', StringType(), True), StructField('Cat_3_level_1', StringType(), True), StructField('Cat_4_level_1', StringType(), True), StructField('Y_From', StructField('Y_From', StringType(), True), StructField('Y_From', St

In []: # Create Hive table with extracted schema
 table_name = "enron"
 schema_ddl = ','.join([f'{field.name} {field.dataType.simpleString()}' for field in schema])
 create_table_query = f"CREATE TABLE IF NOT EXISTS {table_name} ({schema_ddl}) STORED AS PARQUET"
 spark.sql(create_table_query)

24/03/24 11:46:32 WARN SessionState: METASTORE_FILTER_HOOK will be ignored, since hive.security.authorization.manager is set to instance of HiveAuthorizerFactory.

Out[]: DataFrame[]

In []: # Load data into Hive table
 df1.write.mode("overwrite").saveAsTable(table_name)

In []: spark.stop()

5. Performing Hive Queries

```
In [ ]: # Create SparkSession
        spark = SparkSession.builder \
             .appName("Load Hive Table") \
.config("hive.metastore.uris", "thrift://localhost:9083") \
             .enableHiveSupport() \
             .getOrCreate()
        # Load Hive table into DataFrame
        df = spark.sql("SELECT * FROM enron")
        # Show DataFrame
        df.show()
       24/03/24 11:55:41 WARN SparkSession: Using an existing Spark session; only runtime SQL configurations will take effect.
       [Stage 3:>
```

(0 + 1) / 1]

```
From
                                                                                         Subject
                                                                                                              X From
          Message_ID|
                                  Date|
                                                                                                                                   X_To
            X_Folder|
                                X_Origin|
                                                 X_FileName|
                                                                        content|
                                                                                                         Cat_1_level_1|
                                                                                                                             Cat_1_level_2|
                                                                                                                                                 Cat_1_weight|
                                                                                                                                                                   Cat_2_leve
c |
                                                                                              user
          Cat_2_level_2|
                              Cat_2_weight|
                                                 Cat_3_level_1|
                                                                    Cat_3_level_2|
                                                                                        Cat_3_weight|
                                                                                                           Cat_4_level_1|
                                                                                                                              Cat_4_level_2|
                                                                                                                                                  Cat_4_weight|
                                                                                                                                                                     Cat_5_le
l_1|
vel_1|
            Cat_5_level_2|
                                Cat_5_weight|
                                                  Cat_6_level_1|
                                                                      Cat 6 level 2|
                                                                                          Cat_6_weight|
                                                                                                            Cat_7_level_1|
                                                                                                                                Cat 7 level 2|
                                                                                                                                                    Cat_7_weight|
                                                                                                                                                                       Cat_8_
level_1|
             Cat_8_level_2|
                                  Cat_8_weight|
                                                    Cat_9_level_1|
                                                                        Cat_9_level_2|
                                                                                            Cat_9_weight|
                                                                                                             Cat_10_level_1|
                                                                                                                                 Cat_10_level_2|
                                                                                                                                                     Cat_10_weight|
1_level_1|
              Cat_11_level_2|
                                  Cat_11_weight|
                                                     Cat_12_level_1|
                                                                        Cat_12_level_2|
                                                                                             Cat_12_weight|
                                                                                                                      labeled|
_______
|<25542984.1075854...|2000-04-20 10:11:00|frozenset({'eric....|frozenset({'lwbth...|Fwd: Qualify for ...|</pre>
                                                                                                           Eric Bass|lwbthemarine@bigp...|
                                                  ebass.nsf|"Thought you migh...| 19 Apr 2000 20:5...| queueup Copyrigh...| Inc. 2000)with s...| 19 Apr 2000 >20:...|000+ FREE FREE >
                                  Bass-E|
              000 - $99|999 FREE $14.95 >...|000 $14.95 $14.95...| 7 >days a week. ...| November 18| 1999. Barron's i...| money market funds| mutual funds|
                                  000 shares | >each additional... | Inc. Member NASD... | Inc. and is >not... | just hit the rep... | review or >chang... | Free Email at ht... |
tocks| and >bonds >held...|
                    NULL
                                        NULLI
                                                          NULLI
                                                                              NULLI
                                                                                                  NULLI
                                                                                                                     NULLI
                   NULLI
                                      NULLI
                                                          NULLI
                                                                             NULLI
                                                                                                 NULLI
                                                                                                                    NULLI
|<5541885.10758546...|2000-04-20 16:27:00|frozenset({'eric....|frozenset({'timot...|}</pre>
                                                                                                           Eric Bass| Timothy Blanchard|
                                                                                            NULL
                                                                                                                                                      NULL
                                                                                                                                                                          NUL
L|\Eric_Bass_Dec200...|
                                  Bass-E|
                                                  ebass.nsf|-----|
                                                                                            bass-e|
                                                                                                                                                        NULL
                                                                                                                 NULL|
                  NULL
                                     NULL
                                                         NULL
                                                                                                NULL
                                                                                                                   NULL
                                                                                                                                       NULL
                                                                                                                                                          NULL
ULL
NULL|
                   NULL
                                      NULL
                                                          NULL
                                                                             NULL
                                                                                                 NULL
                                                                                                                    NULL
                                                                                                                                       NULL
                                                                                                                                                           NULL
                                                          NULL
                                                                             NULL
NULL
                   NULL
                                      NULL
                                                                                                 NULL
                                                                                                                    NULL
                                                                                                                                       NULL
                                                                                                                                                           NULL
NULL
                   NULL
                                      NULL
                                                          NULL
                                                                             NULL
                                                                                                 NULL
                                                                                                                   False
|<28915064.1075854...|2000-04-24 08:31:00|frozenset({'eric....|"frozenset({""o'n...| 'david.baumbach@...| 'jeffrey.gossett...| 'amir.ahanchian@...| 'kenneth.shulkla...| 'timothy.blanch
                                                                                                                                   Eric Bass|Brian Hoskins, Lu...|
     'nick.hiemstra@e...| 'matthew.lenhart...| 'bryan.hull@enro...| 'luis.mena@enron...| 'brian.hoskins@e...|
                                                                                                             Game Tonight|
                   NULL|\Eric_Bass_Dec200...|
                                                        Bass-E|
                                                                         ebass.nsf|We have a game to...|
                                                                                                                  bass-e|
                                                                                                                                       NULL|
NULL
                                                                                                                                                           NULL
NULL
                   NULL
                                                          NULL
                                                                             NULL
                                                                                                                    NULL
                                                                                                                                       NULL
                                                                                                                                                           NULL
                                      NULL
NULL
                   NULL
                                      NULL
                                                          NULL
                                                                             NULL
                                                                                                 NULL
                                                                                                                    NULL
                                                                                                                                        NULL
                                                                                                                                                           NULL
NULLI
                   NULLI
                                      NULLI
                                                          NULLI
                                                                             NULLI
                                                                                                 NULL
                                                                                                                    NULL
|<6649833.10758546...|2000-04-26 08:55:00|frozenset({'eric....|frozenset({'georg...|}</pre>
                                                                                         WH DATA|
                                                                                                                        George Weissman
                                                                                                           Eric Bass|
                                                                                                                                                      NULL
                                                                                                                                                                          NUL
                                                  ebass.nsf|Here is the wh da...|
L|\Eric_Bass_Dec200...|
                                  Bass-E|
                                                                                            bass-e|
                                                                                                                 NULL
                                                                                                                                     NULL
                                                                                                                                                        NULL
                                     NULL
                                                         NULL
                                                                                                NULL
                                                                                                                   NULL
                                                                                                                                      NULL
                                                                                                                                                          NULL
ULL
                  NULL
                                                         NULL
NULL|
                   NULL
                                      NULL
                                                                             NULL
                                                                                                NULL
                                                                                                                    NULL
                                                                                                                                       NULL
                                                                                                                                                           NULL
                                                                                                                    NULL
NULL
                   NULL
                                      NULL
                                                          NULL
                                                                             NULL
                                                                                                 NULL
                                                                                                                                       NULL
                                                                                                                                                           NULL|
                   NULL
                                      NULL
                                                          NULL
                                                                             NULL
                                                                                                 NULL
                                                                                                                   False|
NULL
|<27501004.1075854...|2000-04-26 11:12:00|frozenset({'eric....|frozenset({'timot...| I'm Having a Party!|</pre>
                                                                                                           Eric Bass|Timothy Blanchard...|
                                                                                                                                                      NULL
             k/E...| dwighthelms@aol.com| Michael Stockton...| Janel Guerrero/C...| April 29th- All ...| Duck Soup will b...| the more the mer...| please call. If ...| let me know by F...|
                    NULLI
                                        NULLI
                                                           NULLI
                                                                               NULL
                                                                                                  NULLI
                                                                                                                      NULLI
bass-el
                   NULLI
                                      NULL
                                                          NULLI
                                                                             NULLI
                                                                                                 NULLI
|<30641730.1075854...|2000-04-26 16:46:00|frozenset({'eric....|frozenset({'lwbth...| Fw: Corruption Test|</pre>
                                                                                                           Eric Bass|"lwbthemarine@big...| mballases@hotmai...| Jason.Bass2@COM
                   NULL
                                      NULL|\Eric_Bass_Dec200...|
                                                                           Bass-E|
                                                                                            ebass.nsf|
                                                                                                                                      bass-e|
                                                                                                                                                           NULL
NULL
                   NULL
                                      NULL
                                                                             NULL
                                                                                                 NULL
                                                                                                                    NULL
                                                                                                                                        NULL
                                                                                                                                                           NULL
                                                          NULL
NULL
                   NULL
                                      NULL
                                                          NULL
                                                                             NULL
                                                                                                 NULL
                                                                                                                    NULL
                                                                                                                                       NULL
                                                                                                                                                           NULL
                   NULL
                                      NULL
                                                          NULL
                                                                             NULL
                                                                                                 NULL
                                                                                                                    NULL
                                                                                                                                       NULL|
                                                                                                                                                           NULL
NULL
                                      NULL|
                                                          NULL
                   NULL
                                                                             NULL
                                                                                                 NULL
                                                                                                                    NULL
                                                                                                           Eric Bass| Shusser@enron.com|
|<2149245.10758546...|2000-04-26 16:47:00|frozenset({'eric....|frozenset({'shuss...| Fw: Corruption Test|</pre>
                                                                                                                                                      NULL
                                                                                                                                                                          NUL
L|\Eric_Bass_Dec200...|
                                  Bass-E|
                                                  ebass.nsf|i bet you score o...|
                                                                                            bass-e|
                                                                                                                 NULL
                                                                                                                                                                           Ν
                                                                                                                                     NULL
                                                                                                                                                        NULL
                                     NULL
                                                                                                                   NULL
                                                                                                                                      NULL
ULL
                  NULL
                                                         NULL
                                                                            NULL
                                                                                                NULL
                                                                                                                                                          NULL
NULL
                                                          NULLI
                                                                             NULL
                                                                                                                                       NULL
                   NULLI
                                      NULL
                                                                                                 NULL
                                                                                                                    NULL
                                                                                                                                                           NULL
NULLI
                   NULLI
                                      NULLI
                                                          NULLI
                                                                             NULLI
                                                                                                 NULLI
                                                                                                                    NULL
                                                                                                                                        NULLI
                                                                                                                                                           NULLI
NULL
                   NULL
                                      NULL
                                                          NULL
                                                                             NULLI
                                                                                                 NULL
                                                                                                                   False|
|<29739912.1075854...|1999-12-14 09:40:00|frozenset({'eric....|frozenset({'danie...|Fw: FROGAPULT, EL...|
                                                                                                           Eric Bass|danielles@jonesgr...|
                                                                                                                                                                          NUL
                                                  ELFBOWL| Y2KGAME Virus Ho...|
                                                                                                                                                                         ELFB
L|\Eric_Bass_Dec200...|
OWL| Y2KGAME Virus Ho...|
                               December 13 | 1999 11:37 AM Su...
                                                                         ELFBOWL| Y2KGAME Virus Ho...| Liesha ----Orig...|
                                                                                                                                December 13 | 1999 10:50 AM To...
                                    ELFBOWL| Y2KGAME Virus Ho...|
                                                                                                                                        NULL
FBOWL| Y2KGAME Virus Ho...|
                                                                            bass-e|
                                                                                                 NULLI
                                                                                                                     NULL
                                                                                                                                                            NULL
NULL
                   NULL
                                      NULL
                                                          NULL
                                                                             NULL|
                                                                                                 NULL
                                                                                                                    NULL
                                                                                                                                        NULL
                                                                                                                                                           NULL
                                                                             NULL
NULL
                                      NULL
                                                          NULL
                                                                                                 NULL
                                                                                                                    NULL
                   NULL
|<17794525.1075854...|2000-04-26 16:47:00|frozenset({'eric....|"frozenset({""o'n...| 'david.baumbach@...| 'bryan.hull@enro...| 'michael.walters...| Fw: Corruption Test|
                                                                                                                                                                     Eric Bas
s|Bryan Hull, O'Nea...|
                                    NULL
                                                       NULL|\Eric_Bass_Dec200...|
                                                                                            Bass-E|
                                                                                                            ebass.nsf|----
                                                                                                                                                      bass-e|
ULL
                  NULL
                                     NULL
                                                         NULL
                                                                            NULL|
                                                                                                NULL
                                                                                                                   NULL
                                                                                                                                      NULL|
                                                                                                                                                          NULL
NULL
                  NULL
                                     NULL
                                                         NULL
                                                                            NULL
                                                                                                NULL
                                                                                                                    NULL
                                                                                                                                       NULL
                                                                                                                                                           NULL
                                      NULL|
                                                                             NULL
                                                                                                                    NULL|
                                                                                                                                        NULL|
NULL
                   NULL
                                                          NULL
                                                                                                 NULL
                                                                                                                                                           NULL
                   NULL
                                      NULL|
                                                          NULL
                                                                             NULL
                                                                                                 NULL
                                                                                                                    NULL
NULL
|<1944629.10758546...|2000-04-27 07:38:00|frozenset({'eric....|frozenset({'timot...| Fw: Corruption Test|</pre>
                                                                                                           Eric Bass| Timothy Blanchard|
                                                                                                                                                      NULL
                                                                                                                                                                          NUL
                                                   ebass.nsf|--
L|\Eric_Bass_Dec200...|
                                  Bass-E|
                                                                                            bass-e|
                                                                                                                 NULL
                                                                                                                                     NULL
                                                                                                                                                        NULL
ULL|
                  NULL
                                      NULL
                                                         NULL
                                                                            NULL
                                                                                                NULL
                                                                                                                   NULL
                                                                                                                                       NULL
                                                                                                                                                          NULL
NULL|
                                      NULL
                                                          NULL|
                                                                             NULL
                                                                                                 NULL
                                                                                                                    NULL
                                                                                                                                       NULL
                                                                                                                                                           NULL
                   NULL
NULL
                                      NULL|
                                                          NULL|
                                                                             NULL
                                                                                                                                        NULL
                                                                                                                                                           NULL
                   NULL
                                                                                                 NULL
                                                                                                                    NULL
NULL
                   NULL
                                      NULL
                                                          NULL
                                                                             NULL
                                                                                                 NULL
                                                                                                                   False
                                                                                                                          Phillip M Love|
                                                                                                                                                                          NUL
|<1373091.10758546...|2000-04-28 07:38:00|frozenset({'eric....|frozenset({'phill...| Fw: Corruption Test|</pre>
                                                                                                           Eric Bass|
                                                                                                                                                      NULL
L|\Eric_Bass_Dec200...|
                                  Bass-E|
                                                  ebass.nsf|---
                                                                  -----|
                                                                                            bass-e|
                                                                                                                 NULL|
                                                                                                                                     NULL
                                                                                                                                                        NULL
                                                         NULL
ULL
                                     NULL
                                                                            NULL
                                                                                                NULL
                                                                                                                   NULL|
                                                                                                                                       NULL
                                                                                                                                                          NULL
                  NULL
NULL
                   NULL
                                      NULL
                                                          NULL
                                                                             NULL
                                                                                                 NULL
                                                                                                                    NULL
                                                                                                                                        NULL
                                                                                                                                                           NULL
NULLI
                   NULL
                                      NULL
                                                          NULL
                                                                             NULL
                                                                                                 NULL
                                                                                                                    NULL
                                                                                                                                        NULL
                                                                                                                                                           NULL
NULL
                   NULL
                                      NULL
                                                          NULL
                                                                             NULL
                                                                                                 NULL
                                                                                                                   False
|<25662208.1075854...|2000-05-01 07:53:00|frozenset({'eric....|"frozenset({""o'n...| 'david.baumbach@...| 'jeffrey.gossett...| 'amir.ahanchian@...| 'kenneth.shulkla...| 'timothy.blanch
     'nick.hiemstra@e...| 'matthew.lenhart...| 'bryan.hull@enro...| 'luis.mena@enron...| 'brian.hoskins@e...| GAME TONIGHT @ 7:45|
                                                                                                                                   Eric Bass|Brian Hoskins, Lu...|
                   NULL|\Eric Bass Dec200...|
                                                                         ebass.nsf|If the weather ho...|
                                                                                                                                        NULL
NULLI
                                                        Bass-El
                                                                                                                  bass-e|
                                                                                                                                                           NULLI
NULL
                   NULL
                                                          NULLI
                                                                             NULLI
                                                                                                                                        NULL
                                                                                                                                                           NULL
                                      NULL
                                                                                                 NULL
                                                                                                                    NULL
NULL
                   NULL
                                      NULL|
                                                          NULL
                                                                             NULL
                                                                                                                                        NULL|
                                                                                                                                                           NULL|
                                                                                                 NULL
                                                                                                                    NULL
NULL
                   NULL
                                      NULL
                                                          NULL
                                                                             NULL
                                                                                                 NULL
                                                                                                                    NULL
|<29617661.1075854...|2000-05-01 09:02:00|frozenset({'eric....|frozenset({'dale....|</pre>
                                                                                      New Product|
                                                                                                           Eric Bass|
                                                                                                                             Dale Neuner | Melba Lozano, Tod... |
                                                                                                                                                                          NUL
                                                                                                                                                        NULL
L|\Eric_Bass_Dec200...|
                                                   ebass.nsf|Dale, On the new ...|
                                                                                                                 NULL
                                                                                                                                     NULL
                                  Bass-E|
                                                                                            bass-e|
                                     NULL
                                                         NULL
                                                                                                                                       NULL
ULL
                  NULL
                                                                            NULL
                                                                                                NULL
                                                                                                                   NULL
                                                                                                                                                          NULL
NULL
                                                          NULL|
                                                                                                                                       NULL
                   NULL
                                      NULL
                                                                             NULL
                                                                                                 NULL
                                                                                                                    NULL
                                                                                                                                                           NULL
NULL
                   NULL
                                      NULLI
                                                          NULLI
                                                                             NULL
                                                                                                 NULL
                                                                                                                    NULL
                                                                                                                                       NULL
                                                                                                                                                           NULL
NULLI
                   NULL
                                      NULLI
                                                          NULL
                                                                             NULLI
                                                                                                 NULL
                                                                                                                   False|
|<4041756.10758546...|2000-05-01 13:40:00|frozenset({'eric....|"frozenset({""o'n...| 'bryan.hull@enro...|Re: Wellhead Accr...|
                                                                                                                              Eric Bass|Bryan Hull, O'Nea...|
                                                                      ebass.nsf|"Here are the num...|000 of accrual va...| accrual term deals| actual volumes g...| half cent bid t
                NULL|\Eric_Bass_Dec200...|
                                                     Bass-E|
o...| incorrect markings| and value withhe...| Greg Sharp will ...| actual volumes g...| and deals incorr...|504 Greg's Value ...|750 Bid to Mid = ...|000 Accrual Value...|000 Accrual Co
sts...|000 Actual Volume...|
                                    000 $118|000 Incorrect Mar...|000) Total Value ...|750 February- Tot...|305 Greg's Value ...|000 Bid to Mid = ...|000 Accrual Value...|000 Accrual
Costs...|000 Actual Volume...|
                                      000 $113|000 Incorrect Mar...|000) Total Value ...|000 March - Total...|145 Greg's Value ...|000 Bid to Mid = ...|000 Accrual Value...|000 Accrual
                                        000 $119|000 Incorrect Mar...|000) Total Value ...|000 Let me know i...| Here are the acc...|
l Costs...|000 Actual Volume...|
|<32439370.1075854...|2000-10-24 09:57:00|frozenset({'eric....|frozenset({'shann...|Re: It could happ...|</pre>
                                                                                                           Eric Bass|
                                                                                                                           Shanna Husser
                                                                                                                                                      NULL
                                                                                                                                                                          NUL
                                                                                            bass-e|
L|\Eric_Bass_Jun200...|
                                                  ebass.nsf|"Thought you migh...|
                                                                                                                                                        NULL|
                                                                                                                                                                           Ν
                                  Bass-E|
                                                                                                                 NULL
                                                                                                                                     NULL
                  NULL
                                     NULL
                                                         NULL
                                                                                                NULL
                                                                                                                   NULL
                                                                                                                                       NULL
                                                                                                                                                          NULL
ULL
                                                                            NULL
NULL
                   NULL
                                      NULL
                                                          NULL
                                                                             NULL
                                                                                                 NULL
                                                                                                                    NULL
                                                                                                                                        NULL
                                                                                                                                                           NULL
NULL
                   NULL
                                      NULL
                                                          NULL
                                                                             NULL
                                                                                                 NULL
                                                                                                                                        NULL
                                                                                                                                                           NULL
                                                                                                                    NULL
                                                          NULL
                                                                             NULL
                                                                                                 NULL
NULL
                   NULL
                                      NULL
                                                                                                                   False|
|<15626432.1075854...|2000-05-01 17:13:00|frozenset({'eric....|frozenset({'timot...|Fwd: Fw: Somethin...|</pre>
                                                                                                           Eric Bass|Brian Hoskins, Ja...|
                                                                                                                                                      NULL
                                                                                                      CHMARGAU@aol.com|
                                                                                                                            srhea@acm.org| Eric Bass/HOU/EC...| ""MONTY WEEKS""
L|\Eric Bass Dec200...|
                                  Bass-E|
                                                  ebass.nsf|"----- John9375@aol.com|
                                                                         ""Short| Michael J"" <bec...| >""Dwayne Shellh...| ""John Schofield...| ""Matthew Rub"" ...| ""RD"" <sabbat
...| ""John Steiert""...|
                                 ""Spillar| >Steve"" <mrmax4...|
                                   ""Palumbo| Donald"" <hardhe...| ""Barry >Mullin"...| ""Gary Monday"" ...| ""Shane >Miller"...| ""Mike McBride""...| >""Doug Marshall...| ""Brett Lawl
h@...| ""Cecil Poe"" ><...|
                   ""LaHue| Robert"" <RLaHue...| ""JettyCat"" ><f...| ""Vernon Hooks""...| ""Aaron >Hommel"...| ""Mike Henderson...| ""John >Handley"...| ""Mike Greene"" ...| >""Roy V.
Clark"...| ""Allan Carruthe...| >""Saul Carrillo...| ""Jeff Campbell"...| ""BREW"" <BREWMA...| ""Johnny Breed""...| ""BeachBum"" <Be...|
|<16080707.1075854...|2000-10-24 10:22:00|frozenset({'eric....|frozenset({'shann...|Re: It could happ...|
                                                                                                          Eric Bass|
                                                                                                                           Shanna Husserl
                                                  ebass.nsf|"----- Kyle Etter/HOU/E...| Kam Keiser/HOU/E...| Jay Reitmeyer/H
L|\Eric_Bass_Jun200...|
                                  Bass-E|
O...| Jeff Coates/HOU/...| William Keeney/H...| Jeffrey C Gosset...| John King/HOU/EC...| Luis Mena/NA/Enr...|
                                                                                                                 @ ENRON| Lisa Gillette/HO...| Susan M Scott/HO...| Dawn C Kenne/
HOU...| Nick Hiemstra/HO...| Benjamin Thomaso...| David Marks/HOU/...| Timothy Blanchar...| you still have t...| none of those ar...| LSU to the Cotto...| the Sharpie Inde...| a sweep wou
                                           WRs| and Defense. Bes...| we have turned t...| the SEC is havin...| but which school...| Matthew Lenhart/...| Christopher Coff...| William K
ld ne...
                       RBs l
elly/H0...| Kyle Etter/H0U/E...| Kam Keiser/H0U/E...| Jay Reitmeyer/H0...| Jeff Coates/H0U/...| William Keeney/H...| Jeffrey C Gosset...|
|<2906108.10758546...|2000-05-04 10:29:00|frozenset({'eric....|frozenset({'dale....|
                                                                                            5X24|
                                                                                                           Eric Bass|
                                                                                                                             Dale Neuner|
                                                                                                                                                      NULL
                                                                                                                                                                          NUL
L|\Eric Bass Dec200...|
                                                  ebass.nsf|Any news? we woul...|
                                  Bass-E|
                                                                                            bass-e|
                                                                                                                 NULL
                                                                                                                                     NULL
                                                                                                                                                        NULL
                                                         NULLI
ULL
                  NULL
                                      NULL
                                                                            NULL
                                                                                                NULL
                                                                                                                   NULL
                                                                                                                                       NULL
                                                                                                                                                          NULL
NULL|
                                                          NULL|
                                                                             NULL
                                                                                                                    NULL|
                                                                                                                                       NULL|
                                                                                                                                                           NULL|
                   NULL
                                      NULL|
                                                                                                 NULL
NULL
                   NULL
                                      NULL
                                                          NULL|
                                                                             NULL
                                                                                                 NULL
                                                                                                                    NULL
                                                                                                                                        NULL
                                                                                                                                                           NULL
NULL
                   NULL
                                      NULL|
                                                          NULL
                                                                             NULL
                                                                                                 NULL
                                                                                                                   False|
|<21265183.1075854...|2000-05-08 07:08:00|frozenset({'eric....|"frozenset({""o'n...| 'david.baumbach@...| 'jeffrey.gossett...| 'amir.ahanchian@...| 'kenneth.shulkla...| 'timothy.blanch
a...| 'nick.hiemstra@e...| 'matthew.lenhart...| 'bryan.hull@enro...| 'luis.mena@enron...| 'brian.hoskins@e...| GAME TONIGHT @ 8:45|
                                                                                                                                   Eric Bass|Brian Hoskins, Lu...|
                   NULL|\Eric_Bass_Dec200...|
                                                        Bass-E|
                                                                         ebass.nsf|Please let me kno...|
                                                                                                                                        NULL
NULL
                                                                                                                  bass-e|
                                                                                                                                                           NULL
```

NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	
NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	
NULL	NULL	NULL	NULL	NULL	NULL	NULL	·	'	
<15691897.1	.075854 2000-05-08	11:46:00 frozenset({'	eric frozenset({'e	earl Daily Thro	ughput	Eric Bass	Earl Tisdale	NULL	NUL
L \Eric_Bass	_Dec200	Bass-E	ebass.nsf Earl, James	s McKay	bass-e	NULL	NULL	NULL	N
ULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	
NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	
NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	
NULL	NULL	NULL	NULL	NULL	NULL	False			
+	+		+				+		
-+	+					+			
+				+	+	+	+		
	+			+	+	+-		+	
	+		+	+	+-			+	
	·	+	+	+		-+	+		
only snowing	j top 20 rows								

6. Linking Neo4j to the Underlying Datastore

```
In []: # connect to neo4j
    uri = "bolt://localhost:7687"
    neo4j_user = "neo4j"
    neo4j_password = "password"
    driver = GraphDatabase.driver(uri, auth=("neo4j", "password"))

# write data to neo4j

df.write \
    .format("org.neo4j.spark.DataSource") \
    .option("url", uri) \
    .option("authentication.basic.username", neo4j_user) \
    .option("authentication.basic.password", neo4j_password) \
    .option("labels", ":Enron") \
    .option("node.keys", "id") \
    .mode("Overwrite") \
    .save()
```

CONCLUSION

This experiment provided us with hands-on experience in leveraging Hive for data storage and querying within a distributed environment. Additionally, we gained proficiency in utilizing Spark to ingest data into Hive tables and execute queries. Moreover, we acquired skills in employing Neo4j for conducting graph-based queries on our dataset. Through this experiment, we gained valuable insights into the capabilities of big data analysis and its potential for extracting meaningful insights from vast datasets.