ASSIGNMENT – III

TITLE :
Create a data security model for encrypting and anonymizing sensitive information in a Big Data system.

NAME : Shinde Shubham Dnyandev. **ROLL NO :** 23107121. **BATCH :** B

```
(base) admin1@cvlcomp07:~$ start-dfs.sh
Starting namenodes on [localhost]
Starting datanodes
localhost: datanode is running as process 6658.  Stop it first and ensure /tm
p/hadoop-admin1-datanode.pid file is empty before retry.
Starting secondary namenodes [cvlcomp07]
cvlcomp07: secondarynamenode is running as process 6862.  Stop it first and e
nsure /tmp/hadoop-admin1-secondarynamenode.pid file is empty before retry.
(base) admin1@cvlcomp07:~$ jps
6658 DataNode
7091 ResourceManager
9987 NameNode
7428 NodeManager
10421 Jps
6862 SecondaryNameNode
(base) admin1@cvlcomp07:~$ nano $HADOOP_HOME/etc/hadoop/core-site.xml
(base) admin1@cvlcomp07:~$ stop-dfs.sh
Stopping namenodes on [localhost]
Stopping datanodes
Stopping secondary namenodes [cvlcomp07]
(base) admin1@cvlcomp07:~$ start-dfs.sh
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [cvlcomp07]
(base) admin1@cvlcomp07:~$ hdfs dfs -ls /
(base) admin1@cvlcomp07:~$ hdfs dfs -mkdir /secure_bank_data
(base) admin1@cvlcomp07:~$ pip install pandas cryptography
Requirement already satisfied: pandas in ./anaconda3/lib/python3.9/site-packa
```

Create a code for Encryption System :

[2]:
```
import pandas as pd
from cryptography.fernet import Fernet
import hashlib
import os
```

[5]:
```
df = pd.read_csv(r"C:\Users\shubh\Downloads\Bank Customer Churn Prediction.csv")
df
```

[5]:

	customer_id	credit_score	country	gender	age	tenure	balance	products_number	credit_card	active_member	estimated_salary	churn
0	15634602	619	France	Female	42	2	0.00	1	1	1	101348.88	1
1	15647311	608	Spain	Female	41	1	83807.86	1	0	1	112542.58	0
2	15619304	502	France	Female	42	8	159660.80	3	1	0	113931.57	1
3	15701354	699	France	Female	39	1	0.00	2	0	0	93826.63	0
4	15737888	850	Spain	Female	43	2	125510.82	1	1	1	79084.10	0
...	...	...	...	...	...	...	...	...	...	...	...	...
9995	15606229	771	France	Male	39	5	0.00	2	1	0	96270.64	0
9996	15569892	516	France	Male	35	10	57369.61	1	1	1	101699.77	0
9997	15584532	709	France	Female	36	7	0.00	1	0	1	42085.58	1
9998	15682355	772	Germany	Male	42	3	75075.31	2	1	0	92888.52	1
9999	15628319	792	France	Female	28	4	130142.79	1	1	0	38190.78	0

10000 rows × 12 columns

[6]:
```
key = Fernet.generate_key()
cipher = Fernet(key)
print("AES Encryption Key : ")
key
```

```
AES Encryption Key :
```

[6]:
```
b'KoujuMGDjLizTr3Hg_41giJHTSJz_8QZ8W8gVfp1Xdw='
```

[7]:
```
# AES Encryption for CustomerId
def encrypt_customer_id(value):
    return cipher.encrypt(str(value).encode()).decode()

# SHA-256 Hashing for Balance
def hash_balance(value):
    return hashlib.sha256(str(value).encode()).hexdigest()

# Data Masking for EstimatedSalary
def mask_salary(value):
    value = str(int(float(value)))
    return "XXXXX" + value[-2:]
```

[8]:
```
# Apply Security Techniques
df["customer_id"] = df["customer_id"].apply(encrypt_customer_id)
df["balance"] = df["balance"].apply(hash_balance)
df["estimated_salary"] = df["estimated_salary"].apply(mask_salary)

print("Secured Data:")
df.head()
```

```
Secured Data:
```

[8]:

	customer_id	credit_score	country	gender	age	tenure	balance	products_nu
0	gAAAAABpgNSczpouyc2fuLUjGu_lyKdfbWYM-5hLA8vn6J...	619	France	Female	42	2	8aed642bf5118b9d3c859bd4be35ecac75b6e873cce34e...	
1	gAAAAABpgNSc45iGTPRH8Sf1BY5MPOsPwE7C1XZcnY2lQv...	608	Spain	Female	41	1	c5954bfe1dd28b3196d60db9c91e868954914a2a37810b...	
2	gAAAAABpgNScAlOVrJvMu0BvD9pQKld7xBu1Ryptkaq71b...	502	France	Female	42	8	e5af7c8c23a99127b4c0670a6aeced0f7f916e042a3ca0...	
3	gAAAAABpgNScA6ZNyf5tEpvHsyaJq4EWgJsKq4u0AgxY6k...	699	France	Female	39	1	8aed642bf5118b9d3c859bd4be35ecac75b6e873cce34e...	
4	gAAAAABpgNScORTaA-M7IwWalQ9pNv7pIQMeHro09P1M-8...	850	Spain	Female	43	2	e5413faf86091ac4f45bfadf842e42fded8b5c0ca0347f...	

[9]:
```
# Save Secured Dataset
output_file = "secured_bank_churn_data.csv"
df.to_csv(output_file, index=False)
print("Secured dataset saved locally.")
```

```
Secured dataset saved locally.
```

[10]:
```
# Upload Secured Data to HDFS
os.system("hdfs dfs -put -f secured_bank_churn_data.csv /secure_bank_data/")
print("Dataset uploaded to HDFS.")
```

```
Dataset uploaded to HDFS.
```

```
(base) admin1@cvlcomp07:~$ hdfs dfs -ls /secure_bank_data
Found 1 items
-rw-r--r--   1 admin1 supergroup    2039732 2026-01-14 10:11 /secure_bank_dat
a/secured_bank_churn_data.csv
(base) admin1@cvlcomp07:~$ hdfs dfs -cat /secure_bank_data/secured_bank_churn
_data.csv | head
customer_id,credit_score,country,gender,age,tenure,balance,products_number,cr
edit_card,active_member,estimated_salary,churn
gAAAAABpZx5IrmBN3amlyISYOPcttq7MTH1zUYhYyG6IMvLavD-Epuk3F2xSv2EpCD8xhGG8stVrh
henXX1Pp61fUBOWlAk_Cw==,619,France,Female,42,2,8aed642bf5118b9d3c859bd4be35ec
ac75b6e873cce34e7b6f554b06f75550d7,1,1,1,XXXXX48,1
gAAAAABpZx5I8l4IIJNs6d7h3gt50VNL91zk_dfbSGIeUwXkKt2l9_rSrTb5DIqRPk1il6A7FlsyN
0M4AMZ7gqCTcpcT3c5HSg==,608,Spain,Female,41,1,c5954bfe1dd28b3196d60db9c91e868
954914a2a37810b280171565d95fcf548,1,0,1,XXXXX42,0
gAAAAABpZx5IBFt2Ejgk4wQFhPIDBpmtX-QYiH6XY_OFDgchsFdGKLC8JxjO5GHXGBPMCCF8UuxGD
Xl0HSS0lC8aK1BZeo3FUw==,502,France,Female,42,8,e5af7c8c23a99127b4c0670a6aeced
0f7f916e042a3ca0b297149e39c20e9983,3,1,0,XXXXX31,1
gAAAAABpZx5ItS2BIWlhdHN8RmRBFSgaCuMneLdAe1pbWgxLSUd_U4N5hSWUM5dH2jRIpyAwXUXjT
v6OMcFv9Jg7-zblqJOZ-g==,699,France,Female,39,1,8aed642bf5118b9d3c859bd4be35ec
ac75b6e873cce34e7b6f554b06f75550d7,2,0,0,XXXXX26,0
gAAAAABpZx5IH1wk1D16CrZTAH0RF2LaWGziPijQ5YbaSVXdIsIAuPpPrRlVhphFAglbNB3eSvvje
fatJxUgibDHAEYIdipMBg==,850,Spain,Female,43,2,e5413faf86091ac4f45bfadf842e42f
```

Output on HDFS :