Steps to be followed:

1. Download File orders.csv from S3 into local (Edgenode)

```
sensor = HttpSensor(
    task_id = 'watch_orders',
    http_conn_id = 'order_s3',
    endpoint='orders.csv',
    response_check = lambda response: response.status_code == 200,
    dag=dag,
    retries = 20,
    retry_delay = timedelta(seconds=20)
)

download_orders = 'rm -rf airflow_pipeline && mkdir -p airflow_pipeline && cd
airflow_pipeline && wget https://trendytech-bigdata.s3.ap-south-
1.amazonaws.com/orders.csv'
```

2. Make corresponding MySQL table on edgenode (orders table)

```
Create table orders(
Order_id INT,
Order_date VARCHAR(255),
Order_customer_id VARCHAR(55),
Order_status VARCHAR(55) );
```

3. Sqoop export from HDFS into MySQL orders table

```
sqoop export \
--connect jdbc:mysql://ms.itversity.com:3306/retail_export \
--username retail_user \
--password itversity \
--table order_airflow \
--export-dir /user/itv003829/airflow/noheader/part* \
--fields-terminated-by ','
```

4. Upload S3 orders.csv into Hive, create Hive table and carryout Sqoop Import into Hive table

```
create table orders1(
order_id int,
order_date string,
order_customer_id string,
order_status string)
 row format delimited
 fields terminated by ',';


sqoop import \
--connect jdbc:mysql://ms.itversity.com:3306/retail_export \
--username retail_user \
--password itversity \
--table order_airflow \
--num-mappers 1 \
--warehouse-dir  /user/itv003829/imported


load data inpath '/user/itv003829/imported/order_airflow'
```

into table orders1;

5. To find customers with Close order Status use Sqoop import with only customers with Order status as CLOSED

```
sqoop import
--connect jdbc:mysql://ms.itversity.com:3306/retail_export
--username retail_user
--password itversity
--table order_airflow
--where "order_status in ('CLOSED')"
--num-mappers 1
--warehouse-dir /user/itv003829/project360/importedclose
--append
```

6. Make Hive table for Customers, (customers details table already available in MySql), use Sqoop import to carry out this step

```
create table customersfinal (
 customer_id int,
 customer_fname string,
 customer_lname string,
 customer_email string,
 customer_password string,
 customer_street string,
 customer_city string,
 customer_state string,
 customer_zipcode string )
row format delimited
fields terminated by ',';
```

```
sqoop import
--connect jdbc:mysql://ms.itversity.com:3306/retail_db
--username retail_user
--password itversity
--table customers
--num-mappers 1
--warehouse-dir /user/itv003829/project360/importedcustomers
--append
```

```
load data inpath '/user/itv003829/project360/importedclose/order_airflow'
into table customersclosed ;
```

7. Combine orders and customers table to make Fact sheet table which contains details of customers with order status as CLOSED. Make Hive HBase table for this to have low latency data availability which in turn will be given to Analytical team

```
create table hiveHbaseFinal(
customer_id int,
customer_fname string,
customer_lname string,
order_id int,
order_date string)
Stored by 'org.apache.hadoop.hive.hbase.HBaseStorageHandler'
With
SERDEPROPERTIES("hbase.columns.mapping"=":key,(per:customer_fname,per:cutomer_lna
me,per:order_id,per:order_date)")
TBLPROPERTIES("hbase.table.name"="projectfinal");

insert overwrite table hiveHbaseFinal select c.customer_id,
c.customer_fname,
c.customer_lname,
o.order_id,
o.order_date
from customerfinal c JOIN ordersclosed o
ON (c.customer_id = o.customer_id);
```