

Demo sản phẩm

A. Stream Layer

Step 1: Chạy kafka

>>> Chạy apache zookeeper

```
/usr/local/kafka/bin/zookeeper-server-start.sh /usr/local/kafka/config/zookeeper.properties
```

>>> Chạy kafka server

```
/usr/local/kafka/bin/kafka-server-start.sh /usr/local/kafka/config/server.properties
```

>>> Chạy create kafka topic

```
/usr/local/kafka/bin/kafka-topics.sh --create --topic smartphoneTopic --bootstrap-server localhost:9092
```

>>> Chạy kafka producer cho stream layer

```
/usr/local/kafka/bin/kafka-console-producer.sh --topic smartphoneTopic --bootstrap-server localhost:9092
```

>>> Chạy kafka consumer cho stream layer

```
/usr/local/kafka/bin/kafka-console-consumer.sh --topic smartphoneTopic --from-beginning --bootstrap-server localhost:9092
```

>>> Sau khi chạy xong có thể kiểm tra kafka trên Kafka tool Offset Explorer 3.0

>>> Mở kafka tool:

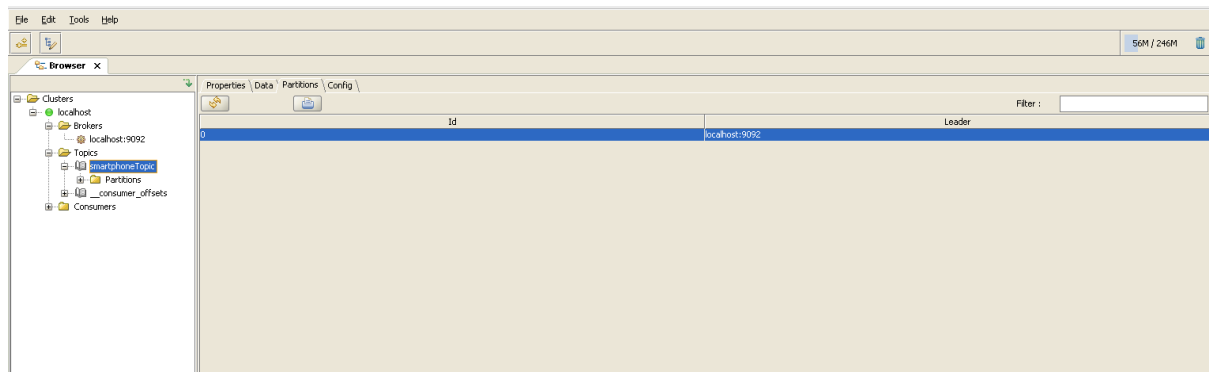
```
cd /
```

```
cd ~/.wine/drive_c/'Program Files'/OffsetExplorer3
```

```
wine offsetexplorer.exe
```

>>> thêm cluster : localhost - 127.0.0.1:9092

>>> Đây là hình ảnh minh họa kafka đã hoạt động, topic “smartphoneTopic” đã được tạo ra, và có 1 partition id = 0 (do cấu hình)



Step 2: start HDFS, YARN

`start-all.sh`

>>> Minh chứng là HDFS, YARN đã chạy

`jps`

```
18753 DataNode
19076 SecondaryNameNode
5093 QuorumPeerMain
20134 Jps
19559 NodeManager
8889 ConsoleProducer
19388 ResourceManager
9692 ConsoleConsumer
5711 Kafka
18575 NameNode
```

>>> Mở web UI của Hadoop

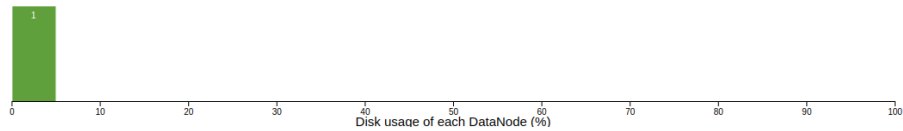
<https://localhost:9870> ⇒ Namenode information

>>> Last Contact (1s), Bộ nhớ chiếm dụng - Used, ..

Datanode Information

✔ In service ❌ Down 🔄 Decommissioning 🛑 Decommissioned 🛑 Decommissioned & dead
🔧 Entering Maintenance 🔧 In Maintenance 🔧 In Maintenance & dead

Datanode usage histogram



In operation

DataNode State: All Show: 25 entries Search:

Node	Http Address	Last contact	Last Block Report	Used	Non DFS Used	Capacity	Blocks	Block pool used	Version
✔ /default-rack/dangkhai:9866 (127.0.0.1:9866)	http://dangkhai:9866	1s	4m	48 KB	39.87 GB	76.35 GB	1	48 KB (0%)	3.3.6

Showing 1 to 1 of 1 entries

Previous 1 Next

Entering Maintenance

No nodes are entering maintenance.

Decommissioning

>>> Thông tin Nodemanager

<https://localhost:8042/>

ResourceManager

NodeManager

Node Information

List of Applications

List of Containers

Tools

NodeManager information

NodeManager info

Total Vmem allocated for Containers	16.80 GB
Vmem enforcement enabled	true
Total Pmem allocated for Container	8 GB
Pmem enforcement enabled	true
Total Vcores allocated for Containers	8
Resource types	memory-mb (unit=M), vcores
NodeHealthyStatus	true
LastNodeHealthTime	Sun May 26 18:50:02 ICT 2024
NodeHealthReport	
NodeManager started on	Sun May 26 17:24:00 ICT 2024
NodeManager Version:	3.3.6 from 1be78238728da9266a4f88195058f08f0d12bf9c by ubuntu source checksum d42eb795a5eadb0febf5e44a7f87a9 on 2023-06-18T08:31Z
Hadoop Version:	3.3.6 from 1be78238728da9266a4f88195058f08f0d12bf9c by ubuntu source checksum 5652179ad5576cb287d9c633bb53bbd on 2023-06-18T08:22Z

Step 3 : Mở hbase, Run thrift server (for Hbase)

/usr/local/hbase/bin/start-hbase.sh

/usr/local/hbase/bin/hbase thrift start

```
khaihadoop@dangkhai:~$ jps
18753 DataNode
19076 SecondaryNameNode
5093 QuorumPeerMain
25478 HMaster
19559 NodeManager
25640 HRegionServer
8889 ConsoleProducer
19388 ResourceManager
9692 ConsoleConsumer
26254 Jps
5711 Kafka
18575 NameNode
```

>>> Web UI của hbase

0.0.0.0:16010

>>> Num region = 2

APACHE

HBASE

[Home](#)[Table Details](#)[Local Logs](#)[Log Level](#)[Debug Dump](#)[Metrics Dump](#)[HBase Configuration](#)

Master dangkhai

Region Servers

Base Stats

Memory

Requests

Storefiles

Compactions

ServerName	Start time	Version	Requests Per Second	Num. Regions
dangkhai,16201,1716719572039	Sun May 26 17:32:52 ICT 2024	1.2.6	0	2
Total:1			0	2

Backup Masters

ServerName	Port	Start Time
Total:0		

Tables

User Tables

System Tables

Snapshots

Tasks

Show All Monitored Tasks

Show non-RPC Tasks

Show All RPC Handler Tasks

Show Active RPC Calls

Show Client Operations

View as JSON

No tasks currently running on this node.

>>> Kiểm tra hbase trên terminal

- Mở hbase
hbase shell
- Kiểm tra có table nào không
list

```
hbase(main):001:0> list
TABLE
0 row(s) in 0.1820 seconds

=> []
hbase(main):002:0>
```

>>> như vậy là chưa có table nào được tạo trên hbase

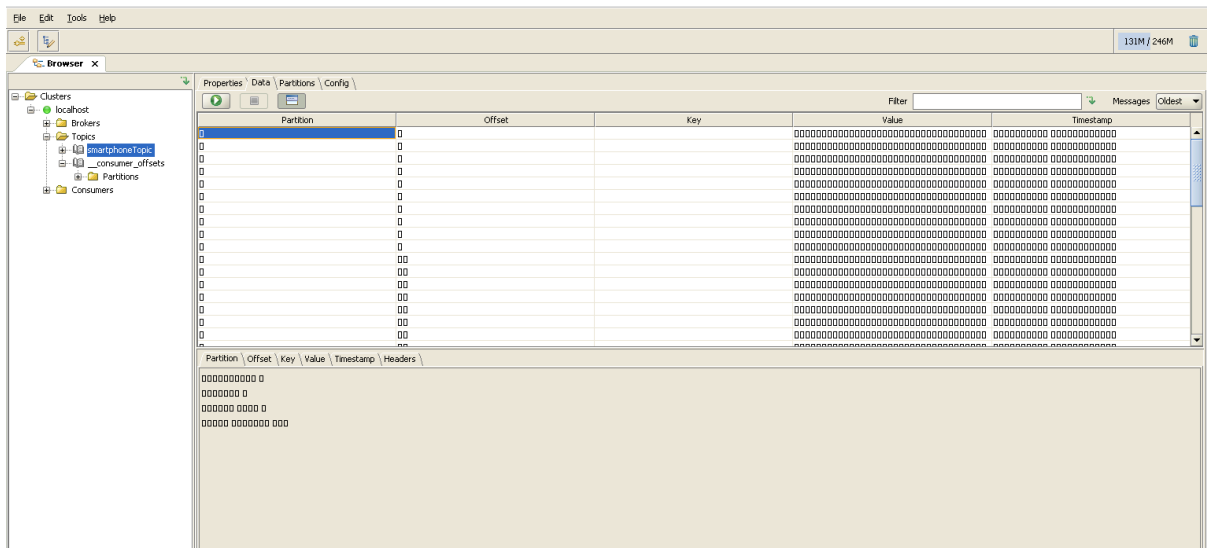
Step 4: Sau tất cả, chạy `stream_pipeline.py` script.

>>> Data đã bắt đầu được đẩy lên “smartphoneTopic” của kafka, tiếp đến là qua Hbase

>>> Đây là terminal của consumer kafka:

```
chaithadoop@dangkhai: $ /usr/local/kafka/bin/kafka-console-consumer.sh --topic smartphoneTopic --from-beginning --bootstrap-server localhost:9092
[{"340", "Infinix", "Smart 7 HD ", "6.6", "2.0", "64.0", "8MP + 5MP", "Dual", "5000.0", "22%", "0.0", "Superfly", "74%", "284", "No reviews"}]
[{"68", "Tecno", "CANON 20 ", "6.67", "8.0", "256.0", "32MP + 64MP + 2MP", "Dual", "5000.0", "28%", "5.0", "SmartNets Enterprises", "74%", "249", "[ (5 out of 5) Good ]"}]
[{"368", "XIAOMI", "Redmi Note 12 ", "6.67", "6.0", "128.0", "50MP + 8MP + 2MP", "Dual", "0.0", "21%", "0.0", "Minestar", "86%", "1002", "No reviews"}]
[{"308", "Tecno", "Pop 6 ", "6.1", "2.0", "32.0", "5MP + 5MP", "Dual", "5000.0", "15%", "5.0", "Little Angels", "40%", "1014", "No reviews"}]
[{"239", "Tecno", "Pop 7 Pro Display ", "6.56", "4.0", "64.0", "13MP + 5MP", "Dual", "5000.0", "24%", "0.0", "Kingsly", "74%", "287", "No reviews"}]
[{"26", "Samsung", "Galaxy A04 ", "6.5", "4.0", "64.0", "50MP + 2MP + 2MP + 5MP", "Dual", "5000.0", "16%", "4.4", "Samsung Ecco World ", "100%", "177", "[ (5 out of 5) I like it ], [ (5 out of 5) Love it ], [ (5 out of 5) Delivery & packaging ], [ (5 out of 5) I just love this phone ], [ (5 out of 5) lovely battery life ], [ (5 out of 5) I like it ], [ (5 out of 5) I love it ], [ (5 out of 5) Best smartphone ], [ (5 out of 5) Samsung A04 ], [ (5 out of 5) Samsung A04 ]"}]
[{"17", "Infinix", "Hot 30i ", "6.6", "8.0", "128.0", "13MP + 8MP", "Dual", "5000.0", "13%", "4.6", "Tech Traders", "86%", "722", "[ (5 out of 5) I like it 100% ], [ (5 out of 5) good battery ], [ (5 out of 5) Good customer service at the pick up session ], [ (5 out of 5) Nice phone ]"}]
[{"361", "Oppo", "Reno10 5G ", "6.7", "8.0", "256.0", "64MP + 32MP + 8MP + 32MP", "Dual", "5000.0", "29%", "0.0", "KM KENYA ELECTRONICS", "80%", "1467", "No reviews"}]
[{"284", "Samsung", "Galaxy A03 Core ", "6.5", "2.0", "32.0", "8MP + 5MP", "Dual", "5000.0", "32%", "3.7", "Little Angels", "40%", "1014", "[ (5 out of 5) I like it ]"}]
[{"131", "Tecno", "Spark 10 ", "6.6", "8.0", "128.0", "50MP + 8MP + 50MP", "Dual", "5000.0", "23%", "4.1", "Minestar", "86%", "1002", "No reviews"}]
[{"175", "XIAOMI", "Redmi A2 ", "6.52", "3.0", "64.0", "8MP + 5MP", "Dual", "5000.0", "16%", "0.0", "XIAOMI KE", "100%", "9268", "No reviews"}]
[{"203", "Tecno", "Spark 10C ", "6.6", "12.0", "128.0", "16MP + 8MP", "Dual", "5000.0", "44%", "5.0", "KM KENYA ELECTRONICS", "80%", "1467", "No reviews"}]
[{"358", "Infinix", "Smart 7 HD ", "6.6", "2.0", "64.0", "8MP + 5MP", "Dual", "5000.0", "31%", "4.0", "Little Angels", "40%", "1014", "[ (5 out of 5) jania ]"}]
[{"264", "Oppo", "A18 ", "6.56", "4.0", "128.0", "8MP + 2MP + 5MP", "Dual", "5000.0", "31%", "0.0", "SmartNets Enterprises", "74%", "249", "No reviews"}]
[{"148", "Oppo", "Reno 10 Pro 5G ", "6.7", "12.0", "256.0", "50MP + 32MP + 8MP + 32MP", "Dual", "4600.0", "14%", "0.0", "CELLINY COLLECTIONS", "74%", "967", "No reviews"}]
[{"365", "Oppo", "Reno 10 5G ", "6.7", "8.0", "256.0", "64MP + 32MP + 8MP + 32MP", "Dual", "5000.0", "33%", "0.0", "Minestar", "86%", "1002", "No reviews"}]
[{"270", "Tecno", "Spark 10C ", "6.6", "8.0", "128.0", "16MP + 2MP", "Dual", "5000.0", "41%", "0.0", "Minestar", "86%", "1002", "No reviews"}]
[{"29", "Infinix", "Hot 30 Play ", "6.82", "8.0", "128.0", "16MP + 8MP", "Dual", "6000.0", "22%", "4.4", "Tech Traders", "86%", "722", "[ (5 out of 5) meets my expectations ], [ (5 out of 5) I like it ], [ (5 out of 5) I LIKE IT ], [ (3 out of 5) I like it ], [ (5 out of 5) i like it ], [ (5 out of 5) Fabulous???????? ], [ (5 out of 5) I LIKE IT ], [ (5 out of 5) great ], [ (3 out of 5) I have mixed feelings. ], [ (5 out of 5) I love it ]"}]
[{"306", "Tecno", "CANON 20 Pro ", "6.67", "8.0", "256.0", "32MP + 64MP", "Dual", "5000.0", "26%", "5.0", "Minestar", "86%", "1002", "No reviews"}]
[{"343", "Apple", "one 14 Pro Max ", "6.7", "6.0", "256.0", "12MP + 12MP", "Single", "4323.0", "32%", "0.0", "Brybell Master Ltd", "50%", "16", "No reviews"}]
[{"274", "Nokia", "C12 ", "6.3", "2.0", "64.0", "8MP + 5MP", "Dual", "3000.0", "21%", "0.0", "SmartNets Enterprises", "74%", "249", "No reviews"}]
[{"362", "Infinix", "NOTE 12 VIP ", "6.67", "8.0", "256.0", "108MP + 13MP + 2MP + 16MP", "Dual", "4500.0", "17%", "0.0", "CELLINY COLLECTIONS", "74%", "967", "No reviews"}]
[{"163", "Nokia", "C12 ", "6.3", "2.0", "64.0", "8MP + 5MP", "Dual", "3000.0", "9%", "4.4", "Superfly", "74%", "284", "[ (5 out of 5) I like the phone. ]"}]
[{"105", "XIAOMI", "Redmi Note 12 ", "6.67", "4.0", "128.0", "50MP + 8MP + 2MP + 13MP", "Dual", "5000.0", "36%", "4.3", "KM KENYA ELECTRONICS", "80%", "1467", "No reviews"}]
[{"217", "Samsung", "Galaxy A04s ", "6.5", "4.0", "64.0", "50MP + 2MP + 2MP + 5MP", "Dual", "5000.0", "20%", "5.0", "Avenue Phones", "100%", "696", "No reviews"}]
```

>>> kafka ở trên tool đã thay đổi, có vẻ bị mã hóa



>>> kiểm tra hbase đã thêm được table “smartphone”

list

```
hbase(main):002:0> list
TABLE
smartphone
1 row(s) in 0.0160 seconds

=> ["smartphone"]
hbase(main):003:0>
```

>>> Table liên tục được cập nhật, trong COLUMN+CELL đã xuất hiện key-value của price ⇒ data đã đi qua thành công model XGBoost

scan “smartphone”

```

hbase(main):003:0> scan "smartphone"
ROW                                COLUMN+CELL
20240526181951901395 column=info:Battery, timestamp=1716722391980, value=5000.0
20240526181951901395 column=info:Brand, timestamp=1716722391980, value=Tecno
20240526181951901395 column=info:Price, timestamp=1716722391980, value=2450.357
7
20240526181951901395 column=info:RAM, timestamp=1716722391980, value=8.0
20240526181951901395 column=info:Screen_size, timestamp=1716722391980, value=6.
67
20240526181951901395 column=info:Sim_type, timestamp=1716722391980, value=Dual
20240526181951901395 column=info:Storage, timestamp=1716722391980, value=256.0
20240526181951901395 column=info:date, timestamp=1716722391980, value=2024-05-2
6 18:19:51
20240526181953255358 column=info:Battery, timestamp=1716722393266, value=0.0
20240526181953255358 column=info:Brand, timestamp=1716722393266, value=XIAOMI
20240526181953255358 column=info:Price, timestamp=1716722393266, value=2120.245
6
20240526181953255358 column=info:RAM, timestamp=1716722393266, value=6.0
20240526181953255358 column=info:Screen_size, timestamp=1716722393266, value=6.
67
20240526181953255358 column=info:Sim_type, timestamp=1716722393266, value=Dual
20240526181953255358 column=info:Storage, timestamp=1716722393266, value=128.0
20240526181953255358 column=info:date, timestamp=1716722393266, value=2024-05-2
6 18:19:53
20240526181958321898 column=info:Battery, timestamp=1716722398333, value=5000.0
20240526181958321898 column=info:Brand, timestamp=1716722398333, value=Tecno
20240526181958321898 column=info:Price, timestamp=1716722398333, value=875.1683
3
20240526181958321898 column=info:RAM, timestamp=1716722398333, value=2.0
20240526181958321898 column=info:Screen_size, timestamp=1716722398333, value=6.
1
20240526181958321898 column=info:Sim_type, timestamp=1716722398333, value=Dual
20240526181958321898 column=info:Storage, timestamp=1716722398333, value=32.0
20240526181958321898 column=info:date, timestamp=1716722398333, value=2024-05-2
6 18:19:58
20240526182003252062 column=info:Battery, timestamp=1716722403270, value=5000.0
20240526182003252062 column=info:Brand, timestamp=1716722403270, value=Tecno
20240526182003252062 column=info:Price, timestamp=1716722403270, value=1185.370
6
20240526182003252062 column=info:RAM, timestamp=1716722403270, value=4.0
20240526182003252062 column=info:Screen_size, timestamp=1716722403270, value=6.
56

```

>>> Trên Web UI (0.0.0.0:16010) của Hbase, số lượng request liên tục thay đổi

Table

smartphone

Table Attributes

Attribute Name	Value	Description
Enabled	true	Is the table enabled
Compaction	NONE	Is the table compacting

Table Regions

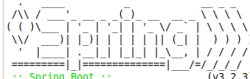
Name	Region Server	Start Key	End Key	Locality	Requests
smartphone_1716722389610.4ff9cfdc0d0f7a7aeb5a001e14c7c9a4.	dangkhai:16301			0.0	259

Regions by Region Server

Region Server	Region Count
dangkhai:16301	1

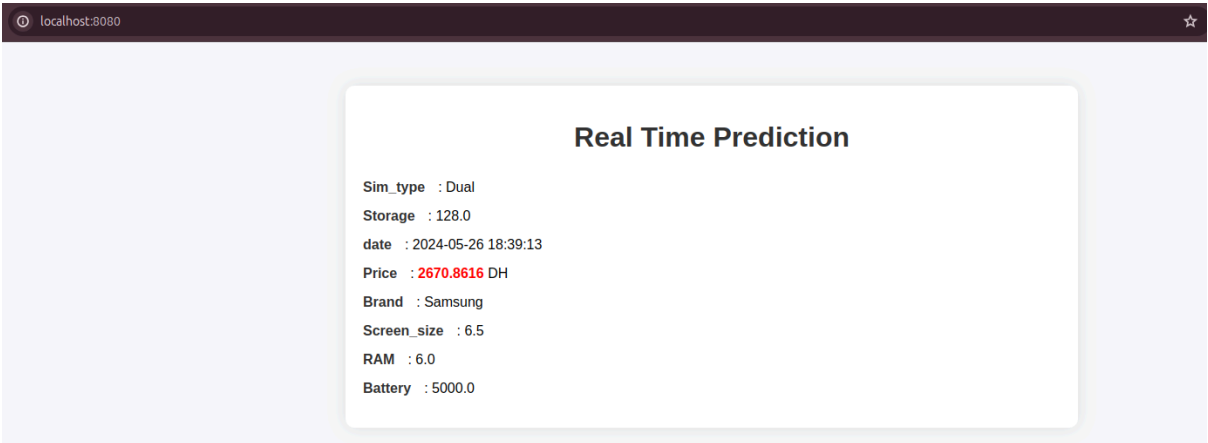
Step 6: Mở ứng dụng Spring Boot Suit 4 để chạy giao diện của lớp Stream layer

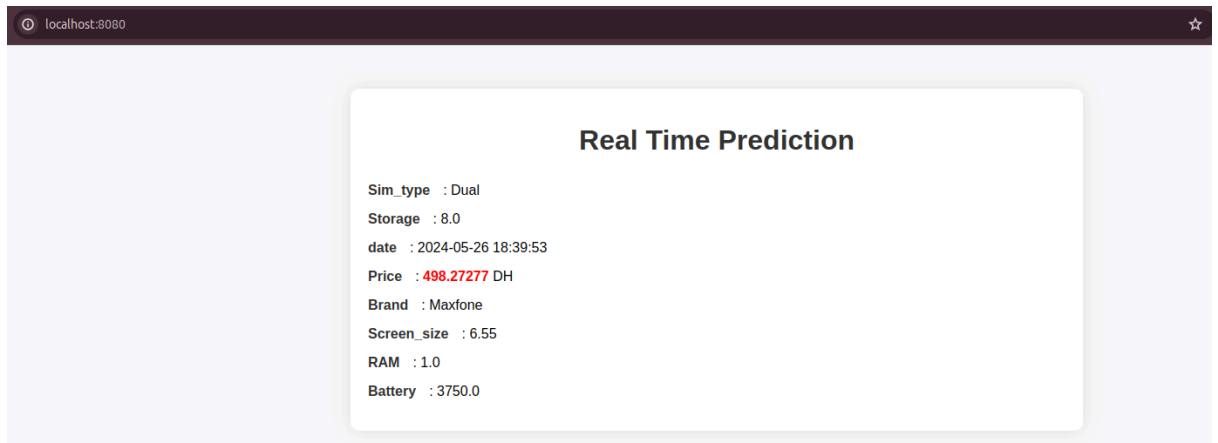
- Mở app spring boot suit
- Run as ‘spring boot app’
- Giao diện spring boot suit



```
2024-05-26T18:36:50.983+07:00 INFO 67721 --- [real_time_app] [main] c.example.demo.RealTimeAppApplication : Starting RealTimeAppApplication using Java 17.0.11 with PID 67721 (/home/khaihad
2024-05-26T18:36:50.989+07:00 INFO 67721 --- [real_time_app] [main] c.example.demo.RealTimeAppApplication : No active profile set, falling back to 1 default profile: "default"
2024-05-26T18:36:52.219+07:00 INFO 67721 --- [real_time_app] [main] o.s.b.w.embedded.tomcat.TomcatWebServer : Tomcat initialized with port 8080 (http)
2024-05-26T18:36:52.238+07:00 INFO 67721 --- [real_time_app] [main] o.apache.catalina.core.StandardService : Starting service [Tomcat]
2024-05-26T18:36:52.239+07:00 INFO 67721 --- [real_time_app] [main] o.apache.catalina.core.StandardEngine : Starting Servlet engine: [Apache Tomcat/10.1.19]
2024-05-26T18:36:52.325+07:00 INFO 67721 --- [real_time_app] [main] o.a.c.c.C.[Tomcat].[localhost].[/] : Initializing Spring embedded WebApplicationContext
2024-05-26T18:36:52.326+07:00 INFO 67721 --- [real_time_app] [main] w.s.c.ServletWebServerApplicationContext : Root WebApplicationContext: initialization completed in 1269 ms
2024-05-26T18:36:52.510+07:00 INFO 67721 --- [real_time_app] [main] o.s.b.a.w.s.WelcomePageHandlerMapping : Adding welcome page template: index
2024-05-26T18:36:52.858+07:00 INFO 67721 --- [real_time_app] [main] o.s.b.w.embedded.tomcat.TomcatWebServer : Tomcat started on port 8080 (http) with context path ''
2024-05-26T18:36:52.870+07:00 INFO 67721 --- [real_time_app] [main] c.example.demo.RealTimeAppApplication : Started RealTimeAppApplication in 2.365 seconds (process running for 3.599)
```

>>> Mở <https://localhost:8080>

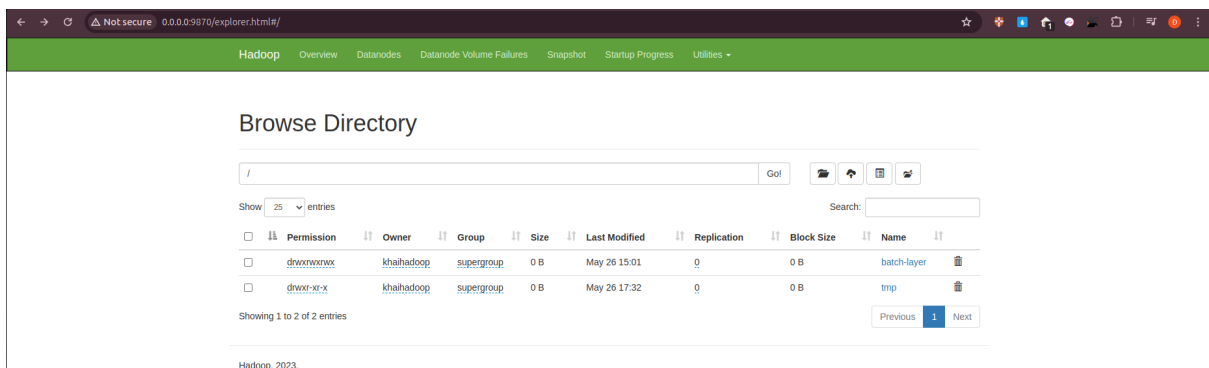




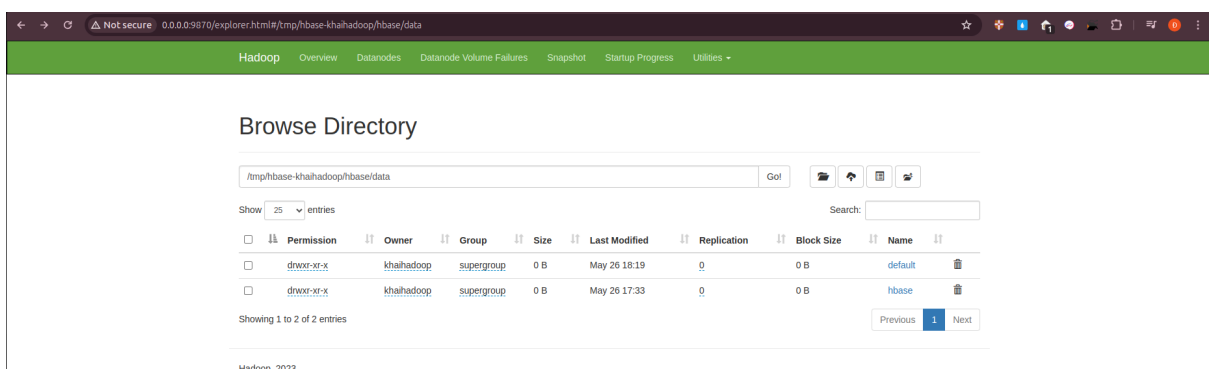
⇒ OK, data đã xuất hiện liên tục theo thời gian (real-time)

>>> Data steam đã được lưu trên hdfs trong thu mục /tmp

<https://localhost:9870/>



>>> vào /tmp/hbase-khaihadoop/hbase/data/



B. Batch-layer

Step 1: Chắc chắn là postgresQL đang chạy

```
sudo service postgresql start
```

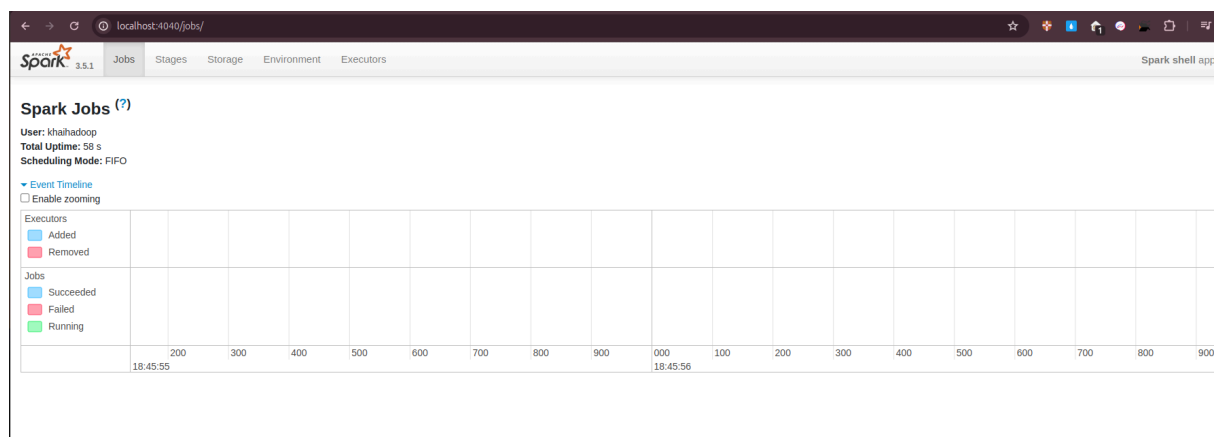
```
khaihadoop@dangkhai:~$ jps
65246 Jps
khaihadoop@dangkhai:~$ sudo service postgresql start
[sudo] password for khaihadoop:
khaihadoop@dangkhai:~$ S
```

Step 2 : start apache spark

```
spark-shell
```

>>> Mở <https://localhost:4040/>

>>> kiểm tra



Step 3: chạy kafka, mở ra producers, consumers cho batch layer

```
/usr/local/kafka/bin/zookeeper-server-start.sh /usr/local/kafka/config/zookeeper.properties
```

```
/usr/local/kafka/bin/kafka-server-start.sh /usr/local/kafka/config/server.properties
```

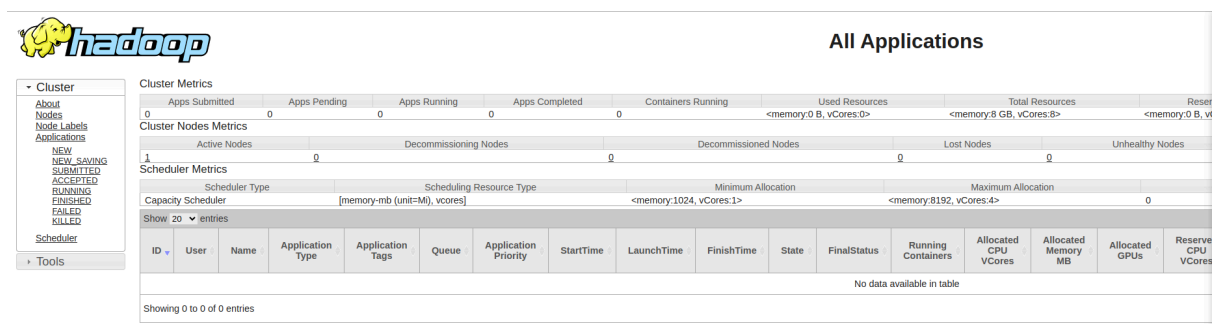
```
/usr/local/kafka/bin/kafka-console-producer.sh --topic smartphoneTopic --bootstrap-server localhost:9092
```

```
/usr/local/kafka/bin/kafka-console-consumer.sh --topic smartphoneTopic --from-beginning --bootstrap-server localhost:9092
```

Step 4 : chạy hdfs, yarn

```
start-all.sh
```

>>> cluster: <http://127.0.0.1:8088/> ⇒ chưa có gì



The screenshot shows the Hadoop YARN web interface. On the left is a navigation menu with options like Cluster, About, Nodes, Node Labels, Applications, and Scheduler. The main area is titled 'All Applications' and displays various metrics and a table of applications. The metrics section shows 0 apps submitted, 0 pending, 0 running, and 0 completed. The table below has columns for ID, User, Name, Application Type, Application Tags, Queue, Application Priority, StartTime, LaunchTime, FinishTime, State, FinalStatus, Running Containers, Allocated CPU Vcores, Allocated Memory MB, Allocated GPUs, and Reserve CPU Vcores. The table is currently empty, showing 'No data available in table' and 'Showing 0 to 0 of 0 entries'.

ID	User	Name	Application Type	Application Tags	Queue	Application Priority	StartTime	LaunchTime	FinishTime	State	FinalStatus	Running Containers	Allocated CPU Vcores	Allocated Memory MB	Allocated GPUs	Reserve CPU Vcores
No data available in table																

Step 5 : Trước khi chạy kiểm tra tra Postgresql

>>> command vào terminal của Postgres:

```
sudo -i -u postgres
```

```
psql
```

>>> Kiểm tra xem có database nào

```
\l
```

>>> Ở đây đã có sẵn database “Big-Data-Project” cho dự án này

```
\c Big-Data-Project
```

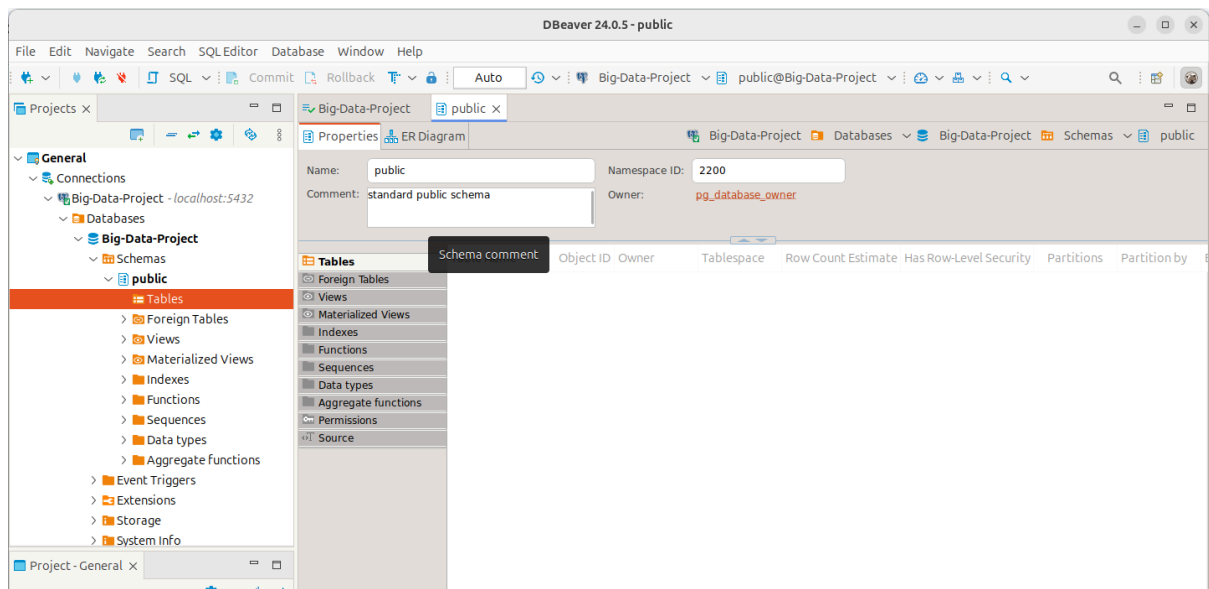
>>> Hiển thị danh sách các bảng trong cơ sở dữ liệu

\dt

>>> không có bảng nào

```
Big-Data-Project=# \dt
Did not find any relations.
Big-Data-Project=#
```

>>> Hoặc có thể kiểm tra trên công cụ DBMS - DBeaver-ce liên kết với database “Big-Data-Project”



>>> Đang không có bảng nào

Step 6: Bắt đầu chạy code

>>> chạy file **batch_pipeline.py** ⇒ chuyển data lên kafka, lưu vào hdfs

```

(5 out of 5) I love it ], [ (5 out of 5) Perfect. ], [ (5 out of 5) I like it ], [ (5 out of 5) Love it ], [ (5 out of 5) A good phone with that price ], [ (3 out of 5) Just an okay phone ], [ (2 out of 5) I don't like it ], [ (5 out of 5) Excellent ], [ (1 out of 5) poor network "] to Kafka topic: smartphoneTopic
Message sent to Kafka topic
Produced: ['1', 'Infinix', 'Hot 30 Play ', '6.82', '8.0', '128.0', '16MP + 8MP', 'Dual', '6000.0', '22%', '4.7', 'Tech Traders', '86%', '722', '[ (5 out of 5) GOOD ], [ (5 out of 5) I like it ], [ (5 out of 5) Like it! ], [ (5 out of 5) phone ], [ (5 out of 5) I love it ], [ (5 out of 5) Awesome! ], [ (4 out of 5) The Phone is great. ], [ (5 out of 5) I like it. ], [ (4 out of 5) Nice camera! However, put up 50MP main camera ], [ (4 out of 5) Like it so far! '] to Kafka topic: smartphoneTopic
Message sent to Kafka topic
Produced: ['246', 'Infinix', 'Smart 7 HD ', '6.6', '2.0', '64.0', '8MP + 5MP', 'Dual', '5000.0', '32%', '5.0', 'Little Angels', '40%', '1014', 'No reviews'] to Kafka topic: smartphoneTopic
Message sent to Kafka topic
Produced: ['344', 'Oppo', 'A58 ', '6.72', '6.0', '128.0', '50MP + 2MP + 8MP', 'Dual', '5000.0', '35%', '0.0', 'Ministar', '86%', '1002', 'No reviews'] to Kafka topic: smartphoneTopic
Message sent to Kafka topic
Produced: ['360', 'Tecno', 'CAMON 20 ', '6.67', '8.0', '256.0', '32MP + 64MP + 2MP', 'Dual', '5000.0', '28%', '0.0', 'SmartNets Enterprises', '74%', '249', 'No reviews'] to Kafka topic: smartphoneTopic
Message sent to Kafka topic
Produced: ['215', 'Samsung', 'Galaxy A14 4G ', '6.6', '4.0', '128.0', '50MP + 2MP + 2MP + 13MP', 'Dual', '5000.0', '14%', '0.0', 'NGALA STORE', '94%', '123', 'No reviews'] to Kafka topic: smartphoneTopic
Message sent to Kafka topic
Produced: ['47', 'Tecno', 'Spark 10 ', '6.6', '8.0', '128.0', '50MP + 8MP + 50MP', 'Dual', '5000.0', '23%', '4.5', 'Ministar', '86%', '1002', '[ (5 out of 5) nice ], [ (5 out of 5) I like it. Thank you!! ], [ (5 out of 5) wonderful. ]'] to Kafka topic: smartphoneTopic
Message sent to Kafka topic
Produced: ['119', 'Tecno', 'Spark 10 Pro ', '6.8', '8.0', '256.0', '50MP + 2MP + 32MP', 'Dual', '5000.0', '38%', '5.0', 'Ministar', '86%', '1002', '[ (5 out of 5) Good. ]'] to Kafka topic: smartphoneTopic
Message sent to Kafka topic
Produced: ['360', 'Tecno', 'CAMON 20 ', '6.67', '8.0', '256.0', '32MP + 64MP + 2MP', 'Dual', '5000.0', '28%', '0.0', 'SmartNets Enterprises', '74%', '249', 'No reviews'] to Kafka topic: smartphoneTopic
Message sent to Kafka topic
Produced: ['304', 'Tecno', 'CAMON 20 Pro ', '6.67', '8.0', '256.0', '32MP + 64MP', 'Dual', '5000.0', '26%', '5.0', 'Ministar', '86%', '1002', 'No reviews'] to Kafka topic: smartphoneTopic
Message sent to Kafka topic
Produced: ['52', 'Infinix', 'Hot 30 Play ', '6.82', '8.0', '128.0', '16MP + 8MP', 'Dual', '6000.0', '22%', '0.0', 'CELLINY COLLECTIONS', '74%', '967', 'No reviews'] to Kafka topic: smartphoneTopic
Message sent to Kafka topic

```

>>> kiểm tra file /batch-layer/raw_data.csv trên hdfs

>>> new terminal

hdfs dfs -cat /batch-layer/raw_data.csv

```

kxathadoop@dangkhat:~$ hdfs dfs -cat /batch-layer/raw_data.csv
,brand,model_name,screen_size,ram,rom,cams,sin_type,battery,sale_percentage,product_rating,seller_name,seller_score,seller_followers,Reviews
0,Tecno,Spark 10 Pro ,6.8,8.0,128.0,50MP + 2MP + 32MP,Dual,5000.0,43%,4.4,Ministar,86%,1002,No reviews
1,Infinix,Hot 30 Play ,6.82,8.0,128.0,16MP + 8MP,Dual,6000.0,22%,4.7,Tech Traders,86%,722,"[ (5 out of 5) GOOD ], [ (5 out of 5) I like it ], [ (5 out of 5) Like it! ], [ (5 out of 5) phone ], [ (5 out of 5) I love it ], [ (5 out of 5) Awesome! ], [ (4 out of 5) The Phone is great. ], [ (5 out of 5) I like it. ], [ (4 out of 5) Nice camera! However, put up 50MP camera ], [ (4 out of 5) Like it so far! ]"
2,XIAOMI,Redmi A2 ,6.52,3.0,64.0,8MP + 5MP,Dual,5000.0,28%,3.8,Xiaomi Ecco World - AC,80%,36,"[ (5 out of 5) I like it ], [ (4 out of 5) I like it ], [ (4 out of 5) I like it ], [ (5 out of 5) fare performance ], [ (5 out of 5) not worthy ], [ (2 out of 5) Slow Charging ], [ (5 out of 5) redmi A2+ ]"
3,Tecno,Spark 10C ,6.6,4.0,128.0,16MP + 8MP,Dual,5000.0,35%,4.0,Ministar,86%,1002,"[ (5 out of 5) It's a good product ], [ (4 out of 5) smart watch ], [ (5 out of 5) Cool ], [ (5 out of 5) product and service delivery ], [ (5 out of 5) I like it ], [ (5 out of 5) Very nice!! ], [ (5 out of 5) Techno spark 10 ]"
4,Infinix,Smart 7 HD ,6.6,2.0,64.0,8MP + 5MP,Dual,5000.0,27%,4.3,Tech Traders,86%,722,"[ (5 out of 5) I like it ], [ (5 out of 5) I like it ], [ (5 out of 5) Good Product ], [ (5 out of 5) I love it ], [ (4 out of 5) I like it ], [ (4 out of 5) Like it ], [ (4 out of 5) Good ], [ (5 out of 5) I like it ], [ (5 out of 5) Battery ], [ (5 out of 5) I love it ], [ (5 out of 5) battery health ], [ (5 out of 5) smartphone ], [ (4 out of 5) She likes it. ], [ (5 out of 5) XIAOMI Redmi A2+ ], [ (5 out of 5) Perfect one ], [ (5 out of 5) I like it ], [ (4 out of 5) redmi A2+ ], [ (5 out of 5) I like it ]"
5,XIAOMI,Redmi A2 ,6.52,2.0,32.0,8MP + 5MP,Dual,5000.0,20%,4.3,XIAOMI KE,100%,9268,"[ (5 out of 5) I love it ], [ (5 out of 5) smartphone ], [ (4 out of 5) She likes it. ], [ (5 out of 5) XIAOMI Redmi A2+ ], [ (5 out of 5) Perfect one ], [ (5 out of 5) I like it ], [ (4 out of 5) redmi A2+ ], [ (5 out of 5) I like it ]"
6,Freeyond,F9 ,6.52,4.0,64.0,3.0MP + 2.0MP + 8.0MP,Dual,5000.0,42%,4.6,FreeYond-COD,86%,443,"[ (5 out of 5) I love the phone ], [ (5 out of 5) i love it ], [ (5 out of 5) well de ], [ (4 out of 5) works well ], [ (4 out of 5) Not bad ], [ (5 out of 5) power ], [ (5 out of 5) good ], [ (5 out of 5) internet ], [ (5 out of 5) i like it ], [ (5 out of 5) to
7,Maxfone,MAX 80 Plus ,6.52,2.0,16.0,5MP + 8MP,Dual,5075.0,0,4.1,Modern Electronics,80%,18,No reviews
8,Infinix,Smart 7 HD ,6.6,2.0,64.0,8MP + 5MP,Dual,5000.0,24%,4.5,Tech Traders,86%,722,"[ (5 out of 5) camera ], [ (3 out of 5) Like it ], [ (5 out of 5) i love it ], [ (4 out of 5) like it ], [ (4 out of 5) nice ], [ (5 out of 5) Excellent ], [ (5 out of 5) got what i ordered ], [ (4 out of 5) i liked it ], [ (5 out of 5) ????????? ], [ (5 out of 5) CONVINIER
9,Freeyond,M5 ,6.52,8.0,128.0,50MP + 8.0MP + 2.0MP,Dual,5000.0,35%,4.1,FreeYond-COD,86%,443,"[ (1 out of 5) product specification ], [ (5 out of 5) I love it ], [ (5 out of 5) Pe ], [ (5 out of 5) I like it ], [ (5 out of 5) Love it ], [ (5 out of 5) A good phone with that price ], [ (3 out of 5) Just an okay phone ], [ (2 out of 5) I don't like it ], [ (5 out of 5) Excellent ], [ (1 out of 5) poor network ]"
10,Tecno,Pop 7 Pro ,6.6,2.0,64.0,5MP + 8MP,Dual,5000.0,19%,4.4,Premium Phones-,100%,407,"[ (5 out of 5) I love it ], [ (5 out of 5) I like it ], [ (5 out of 5) I like it ], [ (4 out of 5) Very good ], [ (5 out of 5) A good device ], [ (4 out of 5) rate ], [ (5 out of 5) good phone ], [ (2 out of 5) good ], [ (4 out of 5) great ], [ (5 out of 5) Tecno Pop 7 ]"
11,Tecno,Pop 7 ,6.6,2.0,64.0,5MP + 8MP,Dual,5000.0,11%,4.4,Premium Phones-,100%,407,"[ (4 out of 5) Delivery process excellent ], [ (5 out of 5) i loved it ], [ (4 out of 5) As bed ], [ (4 out of 5) Nice phone ], [ (5 out of 5) Awesome one ], [ (3 out of 5) Poor camera ], [ (5 out of 5) quality ]"
12,Infinix,Hot 30i ,6.56,4.0,128.0,13MP + 8MP,Dual,5000.0,29%,4.4,Tech Traders,86%,722,"[ (5 out of 5) Infinix hot30i ], [ (1 out of 5) Really poor phone ], [ (5 out of 5) I like ], [ (5 out of 5) I like it ], [ (5 out of 5) Decent ], [ (5 out of 5) original ], [ (5 out of 5) i liked it ], [ (5 out of 5) Love it ], [ (4 out of 5) I like it ], [ (5 out of 5) BUDGET ]"
13,XIAOMI,Redmi A2 ,6.52,2.0,32.0,8MP + 5MP,Dual,5000.0,23%,4.7,XIAOMI KE,100%,9268,"[ (5 out of 5) Fast, efficient and easier to operate. ], [ (5 out of 5) Loved it ], [ (5 out of 5) performance ], [ (5 out of 5) its the best ], [ (5 out of 5) I like it. ], [ (5 out of 5) Perfect! ], [ (5 out of 5) I like it. ], [ (5 out of 5) I love it ], [ (5 out of 5) I love ], [ (5 out of 5) love ]"
14,Tecno,Pop 7 Pro ,6.6,4.0,64.0,13MP + 5MP,Dual,5000.0,17%,4.7,Premium Phones-,100%,407,"[ (5 out of 5) Good ]"
15,Huawei,P Smart ,6.21,6.0,128.0,13MP + 2MP + 8MP,Dual,0.0,44%,0.0,3C Market-COD,60%,22,No reviews
16,Tecno,Spark 10C ,6.6,4.0,128.0,16MP + 8MP,Dual,5000.0,18%,4.1,Premium Phones-,100%,407,"[ (5 out of 5) spark 10c ], [ (4 out of 5) product arrived in good condition ], [ (2 out of 5) Camera so poor ], [ (5 out of 5) very nice phone ], [ (5 out of 5) I like it ], [ (4 out of 5) Great ], [ (5 out of 5) PERFORMANCE ], [ (5 out of 5) Efficiency ], [ (5 out of 5) d good ], [ (5 out of 5) Good specs ]"
17,Infinix,Hot 30i ,6.6,8.0,128.0,13MP + 8MP,Dual,5000.0,13%,4.6,Tech Traders,86%,722,"[ (5 out of 5) I like it 100% ], [ (5 out of 5) good best ], [ (5 out of 5) Good customer s at the pick up session ], [ (5 out of 5) Nice phone ]"

```

>>> chạy file **syc_batch_pipeline.py** để lấy data lưu ở hdfs đi qua spark transformation (đi qua model xgboost) rồi cuối cùng đẩy lên postgres:

https://xgboost.readthedocs.io/en/stable/tutorials/saving_model.html
for more details about differences between saving model and serializing.

```
warnings.warn(msg, UserWarning)
Data transformed successfully
Data stored in PostgreSQL
begin
Data transformed successfully
Data stored in PostgreSQL
begin
Data transformed successfully
Data stored in PostgreSQL
begin
Data transformed successfully
Data stored in PostgreSQL
begin
Data transformed successfully
Data stored in PostgreSQL
begin
Data transformed successfully
```

>>> Kiem tra web Spark

Page: 1

1 Pages. Jump to . Show items in a page.

[illegible]

Page: 1

1 Pages. Jump to 1, Show 100 items in a page, Go

>>> data đã có trên postgresql

The screenshot shows the DBeaver application window with a table named 'Phone' from the 'public' schema. The table contains 42 rows of smartphone data. The columns are: id, brand, model_name, screen_size, ram, rom, and cams. The interface includes a sidebar with 'General' and 'Data' tabs, and a bottom status bar indicating '42 row(s) fetched'.

	id	brand	model_name	screen_size	ram	rom	cams
1	44	6	A78	6.43	8	256	50MP
2	149	7	G11 Plus	6.52	4	64	50MP + 2MP + 8MP
3	324	2	NOTE 30	6.66	8	256	64MP + 2MP
4	51	5	CAMON 20	6.67	8	256	32MP + 64MP + 2MP
5	91	8	Galaxy A14 4G	6.6	4	128	50MP + 2MP + 2MP + 13MP
6	128	4	XIAOMI 10A	6.53	6	128	13MP + 5MP
7	66	6	A58	6.72	6	128	50MP + 2MP + 8MP
8	317	10	Y36	6.64	8	128	50MP + 2MP + 16MP
9	64	12	Find 30	6.6	2	32	5MP + 13MP
10	230	8	Galaxy Note20 Ultra 5G	6.9	12	128	108MP + 12MP + 12MP + 10MP
11	310	8	Galaxy A03 Core	6.5	2	32	8MP + 5MP
12	38	5	Spark 10C	6.6	8	128	16MP + 8MP
13	220	5	Pop 7	6.6	2	64	5MP + 8MP
14	269	11	C33	6.5	4	128	50MP + 5MP
15	101	5	CAMON 20 Pro 4G Mr Doodle Edition	6.67	16	256	32MP + 64MP + 2MP
16	128	4	XIAOMI 10A	6.53	6	128	13MP + 5MP
17	225	6	A58	6.72	8	128	50MP + 2MP + 8MP
18	41	8	Galaxy A14	6.6	4	64	50MP + 5MP + 2MP + 2MP + 13MP
19	90	8	Galaxy F14 5G	6.6	6	128	13MP + 50MP + 2MP + 2MP
20	86	8	Galaxy A04	6.5	4	64	5MP
21	236	11	C33	6.5	4	128	50MP + 5MP
22	37	2	SMART 7 PLUS Type C	6.6	9	64	13MP
23	291	4	Redmi 12	6.79	4	128	50MP + 8MP + 2MP + 8MP
24	213	8	Galaxy M34 5G	6.5	6	128	50MP + 8MP + 2MP + 13MP
25	375	5	Spark 10 Disp	6.6	8	128	50MP + 8MP
26	298	2	NOTE 30	6.66	8	256	64MP + 2MP
27	274	7	C12	6.3	2	64	8MP + 5MP
28	237	5	Spark 10 Pro	6.8	24	128	50MP + 32MP
29	206	4	Redmi 12c	6.71	4	128	50MP + 0.08MP + 5MP
30	194	6	A17K	6.56	3	64	8MP + 2MP + 5MP
31	6	3	F9	6.52	4	64	3.0MP + 2.0MP + 8.0MP
32	127	8	Galaxy A14 4G	6.6	4	128	50MP + 2MP + 2MP + 13MP

>>> đã xuất hiện cột price ⇒ data đã được xử lý, đi qua xgboost

Column Name	#	Data type	Identity
123id	1	int8	
123brand	2	int4	
ABCmodel_name	3	text	
123screen_size	4	float4	
123ram	5	float4	
123rom	6	float4	
ABCamcs	7	text	
123sim_type	8	int4	
123battery	9	float4	
ABCsale_percenta	10	text	
123product_ratin	11	float8	
ABCseller_name	12	text	
ABCseller_score	13	text	
123seller_followe	14	int8	
ABCSeller Reviews	15	text	
ABCSprice	16	text	

