

**Term Project – Auction Items Data Processing Pipeline**

Madhavi Ghanta

College of Science and Technology, Bellevue University

**DSC650-T301: Big Data**

Professor. Nasheb Ismaily

March 03, 2024

## **Introduction**

In this term-end project, several big data technologies were used to create a data engineering pipeline to find the top bidders who should be invited for auction in each category. The project is also aimed to provide analytics of state-wide trends of auctions. There are a number of predefined entities using which the outcome should be done:

- New items ready for auction coming in stream
- New users coming in as stream to register for auction site - Kafka topic bidders
- Active Bids coming as stream
- 1 value object of item category- csv file
- A bidder can be for multiple categories

## **Data set used:**

[Auctions.csv](#)

[Bidders.csv](#)

[Bids.csv](#)

[Category.csv](#)

[https://github.com/madhavig2020/dsc650/blob/main/auction%20data%20set%20-%20auctions\\_list.csv](https://github.com/madhavig2020/dsc650/blob/main/auction%20data%20set%20-%20auctions_list.csv)

## **Goal:**

Based on the aforementioned data sets, a solution should be able to calculate which top 20 bidders performed better historically in a specific Item category, which means inviting them for a new auction, as it is coming in a stream of data.

## **Solution:**

**Criteria for solution stack.** All listed technologies chosen based on the core criteria:

- Truly scalable, distributed and highly efficient in data processing and streaming
- Stability
- Result with a minimum configuration
- Popularity, community, ease of integration and ecosystem
- Personal preference

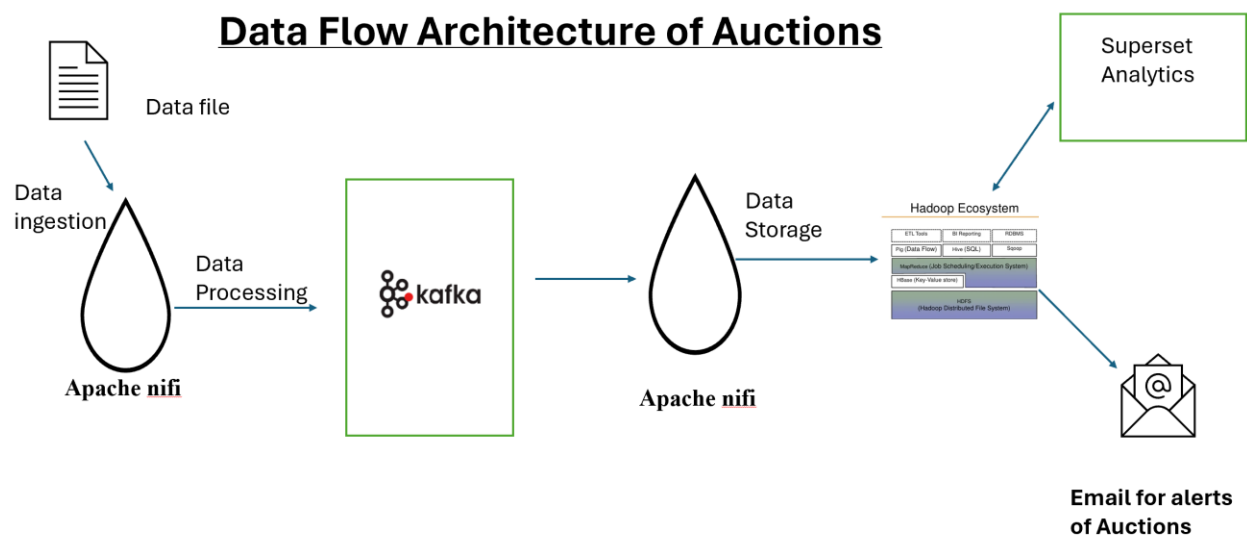
**Apache nifi-** is used for real-time data flow control to help manage the transfer of data between various sources and destination.

**Apache Kafka** - highly scalable, fault-tolerant distributed message streaming / Pub-Sub system. Now #1 on the market.

**Hadoop – HDFS - USES** distributed storage and parallel processing to handle big data and analytics jobs, breaking workloads down into smaller workloads that can be run at the same time.

**Apache Hive** is used to share the results of ranked bidders from Top Category bidders.

## Project Architecture Overview



## Infrastructure Setup

I used Terraform to automate the deployment of the tech stack in GCP. Deployed the docker images given in the course and ensured that they were up and running.

```
ag13@AG13s-MacBook-Pro ~ % ls -l
total 2342528
drwxr-xr-x 15 ag13 staff 480 Feb 26 12:11 Ag13-Data
drwx-----@ 4 ag13 staff 128 Oct 9 19:29 Applications
drwx-----+ 17 ag13 staff 544 Mar 3 15:12 Desktop
drwx-----+ 4 ag13 staff 128 Feb 28 12:08 Documents
drwx-----+ 20 ag13 staff 640 Mar 3 14:45 Downloads
drwx-----@ 87 ag13 staff 2784 Mar 1 01:07 Library
drwx----- 4 ag13 staff 128 Mar 30 2023 Movies
drwx-----+ 4 ag13 staff 128 Sep 2 2023 Music
drwx-----+ 4 ag13 staff 128 Mar 2 08:57 Pictures
drwxr-xr-x+ 4 ag13 staff 128 Feb 24 2023 Public
drwxr-xr-x 6 ag13 staff 192 Mar 3 09:33 dsc650-infra
drwxr-xr-x 2 ag13 staff 64 Feb 26 20:24 key
drwxr-xr-x 7 ag13 staff 224 Feb 26 18:53 learn-terraform-docker-container
drwxr-xr-x 9 ag13 staff 288 Feb 29 07:17 learn-terraform-gcp
drwxr-xr-x 3 ag13 staff 96 Mar 3 12:13 mg-auction
drwxr-xr-x 3 ag13 staff 96 Feb 29 23:31 mg-dsc650-infra-bak
drwxrwxr-x@ 19 ag13 staff 608 Mar 1 01:07 nifi-1.25.0
-rw-r--r--@ 1 ag13 staff 1199370198 Feb 27 21:58 nifi-1.25.0-bin.zip
-rwxrwxrwx 1 ag13 staff 99 Mar 1 11:10 nifi.sh
drwxrwxrwx 3 ag13 staff 96 Mar 3 10:59 test-data
drwxr-xr-x 18 ag13 staff 576 Mar 1 01:07 vehicle-auction-stream-processor
ag13@AG13s-MacBook-Pro ~ %
```

## Terraform Script used:



```
mg-terraform-gcp — vi main.tf — 140x75

Terraform {
  required_providers {
    google = {
      source = "hashicorp/google"
      version = "4.51.0"
    }
  }
}

provider "google" {
  credentials = file("${abspath(path.root)}/mghanta-2024-dsc650-project-359ad868382e.json")

  project = "mghanta-2024-dsc650-project"
  region = "us-central1"
  zone = "us-central1-c"
}

resource "google_compute_subnetwork" "network-with-private-secondary-ip-ranges" {
  name = "mg-subnet"
  ip_cidr_range = "10.2.0.0/16"
  region = "us-west2"
  network = google_compute_network.dsc-proj-network.id
  secondary_ip_range {
    range_name = "mg-ip-range"
    ip_cidr_range = "192.168.10.0/24"
  }
}

resource "google_compute_network" "dsc-proj-network" {
  name = "mg-vpc-network"
  auto_create_subnetworks = false
}

# This code is compatible with Terraform 4.25.0 and versions that are backwards compatible to 4.25.0.

resource "google_compute_instance" "instance-20240227-143538" {
  boot_disk {
    auto_delete = true
    device_name = "mg-instance"

    initialize_params {
      image = "projects/ubuntu-os-cloud/global/images/ubuntu-2004-focal-v20240226"
      size = 60
      type = "pd-balanced"
    }
  }

  mode = "READ_WRITE"
}

can_ip_forward = false
deletion_protection = false
enable_display = false

labels = {
  goog-ec-src = "vm_add-tf"
}

machine_type = "e2-custom-4-8192"

metadata = {
  startup-script = "#! /bin/bash\nmkdir mg\ncd mg\ngit clone https://github.com/bellevue-university/dsc650-infra.git\ncd /mg/dsc650-infra/\nchmod +x setup.sh\n./setup.sh"
}

name = "mg-instance"

network_interface {
  access_config {
    network_tier = "PREMIUM"
  }
}

queue_count = 0
stack_type = "IPV4_ONLY"
subnetwork = "projects/mghanta-2024-dsc650-project/regions/us-west2/subnetworks/default"
```

## Challenges faced:

Due to the challenges faced for testing number of trails / iterations in bringing up and destroying the tech stack in GCP, I have duplicated the same tech stack environment in my local machine.

Issues identified in local:

- Resolved port conflicts while running Hadoop and Kafka at the same time using port 2181
- Opened ports from outside docker images to put files from nifi to HDFS
- Modified docker-compose files to handle hostname issues

Containers

Images

Volumes

Builds

Dev Environments BETA

Docker Scout

Extensions ⋮

+

Add Extensions

Containers

Give feedback 🗨️

Container CPU usage 📊

25.56% / 400% (4 CPUs available)

Container memory usage 📊

6.54GB / 7.48GB

Show charts

🔍

Search

⏏

⏻

Only show running containers

Delete

▶

⏏

⏻

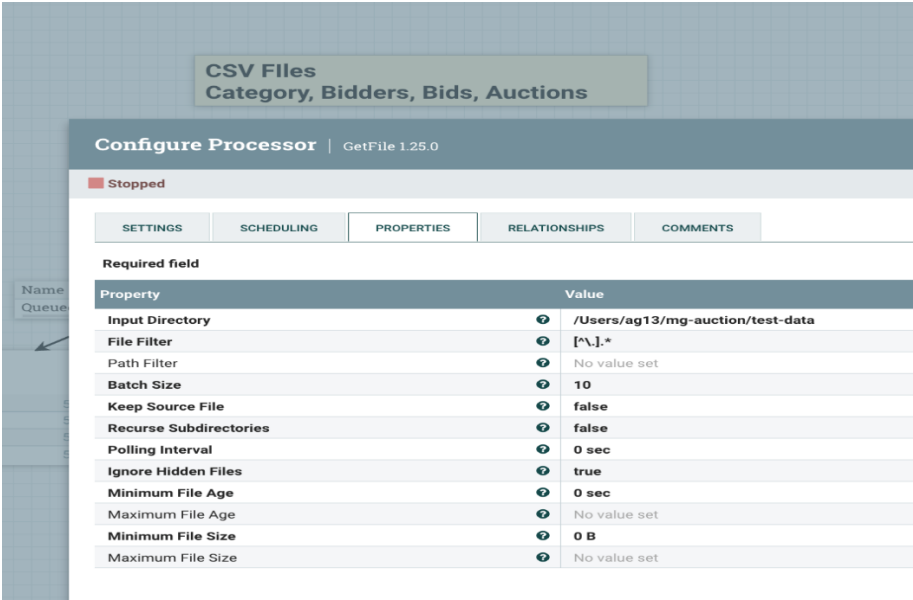
✓	Name ↑	Image	Status	CPU (%)	Port(s)	Last started	Actions
✓	<div>📦</div> <div>hadoop-hive-spark-hbase</div>		Running (7/7)	22.34%		4 hours ago	<div>■</div> <div>⋮</div> <div>🗑️</div>
✓	<div>📦</div> <div>hivemetastore-1</div> <div>517b65cfc048 <span>🗑️</span></div>	postgres:11.5	Running	0.99%	5432:5432 <span>🔗</span>	4 hours ago	<div>■</div> <div>⋮</div> <div>🗑️</div>
✓	<div>📦</div> <div>master-1</div> <div>9220b8c03452 <span>🗑️</span></div>	nasheb/bigdata:1.0	Running	8.76%	10000:10000 <span>🔗</span> Show all ports (15)	4 hours ago	<div>■</div> <div>⋮</div> <div>🗑️</div>
✓	<div>📦</div> <div>worker1-1</div> <div>4f7f4d632e20 <span>🗑️</span></div>	nasheb/bigdata:1.0	Running	5.79%	45081:45080 <span>🔗</span> Show all ports (4)	4 hours ago	<div>■</div> <div>⋮</div> <div>🗑️</div>
✓	<div>📦</div> <div>worker2-1</div> <div>c562b6984d9d <span>🗑️</span></div>	nasheb/bigdata:1.0	Running	5.2%	45082:45080 <span>🔗</span> Show all ports (4)	4 hours ago	<div>■</div> <div>⋮</div> <div>🗑️</div>
✓	<div>📦</div> <div>zoo1-1</div> <div>340b218f0947 <span>🗑️</span></div>	zookeeper	Running	0.53%	2181:2181 <span>🔗</span>	4 hours ago	<div>■</div> <div>⋮</div> <div>🗑️</div>
✓	<div>📦</div> <div>zoo2-1</div> <div>4ee2f753f994 <span>🗑️</span></div>	zookeeper	Running	0.61%	2182:2181 <span>🔗</span>	4 hours ago	<div>■</div> <div>⋮</div> <div>🗑️</div>
✓	<div>📦</div> <div>zoo3-1</div> <div>34b130847966 <span>🗑️</span></div>	zookeeper	Running	0.46%	2183:2181 <span>🔗</span>	4 hours ago	<div>■</div> <div>⋮</div> <div>🗑️</div>
✓	<div>📦</div> <div>kafka</div>		Running (2/2)	1.94%		4 hours ago	<div>■</div> <div>⋮</div> <div>🗑️</div>
✓	<div>📦</div> <div>kafka-kafka-1-1</div> <div>4e31c824f1af <span>🗑️</span></div>	nasheb/kafka:1.0	Running	1.43%	9092:9092 <span>🔗</span>	4 hours ago	<div>■</div> <div>⋮</div> <div>🗑️</div>
✓	<div>📦</div> <div>zookeeper-1</div> <div>aa1a4b61c7cc <span>🗑️</span></div>	nasheb/zookeeper:1.0	Running	0.51%	49579:21814 <span>🔗</span>	4 hours ago	<div>■</div> <div>⋮</div> <div>🗑️</div>
✓	<div>➤</div> <div>solr</div>		Running (1/1)	1.28%		4 hours ago	<div>■</div> <div>⋮</div> <div>🗑️</div>

Selected 12 of 12

Step 1

Data Ingestion – Nifi

Data set- bidders.csv, category.csv , auctions.csv , bids.csv files has been ingested to nifi to docker data mount



Confirmation of data ingestion:



NiFi Data Provenance

Displaying 118 of 118

Oldest event available: 03/01/2024 00:36:29 PST

Filter
by component name

Date/Time	Type
03/03/2024 10:59:11.925 PST	CLON
03/03/2024 10:59:11.925 PST	CLON
03/03/2024 10:59:11.925 PST	CLON
03/03/2024 10:59:11.925 PST	RECE
03/03/2024 10:59:10.923 PST	CLON
03/03/2024 10:59:10.923 PST	CLON
03/03/2024 10:59:10.923 PST	CLON
03/03/2024 10:59:10.923 PST	RECE
03/03/2024 10:59:06.918 PST	CLON
03/03/2024 10:59:06.918 PST	CLON
03/03/2024 10:59:06.918 PST	CLON
03/03/2024 10:59:06.918 PST	RECE
03/03/2024 10:54:42.647 PST	CLON
03/03/2024 10:54:42.647 PST	CLON
03/03/2024 10:54:42.646 PST	CLON
03/03/2024 10:54:42.646 PST	RECE
03/03/2024 10:49:57.190 PST	CLON

### Provenance Event

DETAILS
ATTRIBUTES
CONTENT

#### Input Claim

Container  
No value previously set

Section  
No value previously set

Identifier  
No value previously set

Offset  
No value previously set

Size  
No value previously set

#### Output Claim

Container  
default

Section  
3

Identifier  
1709483775080-3

Offset  
30548

Size  
227 bytes

DOWNLOAD
VIEW

#### Replay

Cannot replay data from Provenance Event because the Source FlowFile Queue with ID No Value no longer exists

Step 2

Data has been mounted to Hadoop docker files.

02

Docker Data mount - PutFile

PutFile 1.25.0

org.apache.nifi - nifi-standard-nar

In0 (0 bytes)

Read/Write0 bytes / 0 bytes

Out0 (0 bytes)

Tasks/Time0 / 00:00:00.000

Configure Processor

PutFile 1.25.0

Stopped

SETTINGS

SCHEDULING

PROPERTIES

RELATIONSHIPS

COMMENTS

Required field

Property	Value
Directory	/Users/ag13/dsc650-infra/bellevue-bigdata/hadoop-hive-spa...
Conflict Resolution Strategy	fail
Create Missing Directories	true
Maximum File Count	No value set
Last Modified Time	No value set
Permissions	777
Owner	No value set
Group	No value set

NiFi Data Provenance

Displaying 46 of 46

Oldest event available: 03/01/2024 00:36:29 PST

Filter

by component name

Date/Time	Type
03/03/2024 14:56:32.138 PST	DROF
03/03/2024 14:56:32.136 PST	DROF
03/03/2024 14:56:32.135 PST	DROF
03/03/2024 14:56:32.133 PST	DROF
03/03/2024 14:56:32.132 PST	DROF
03/03/2024 14:56:32.131 PST	SEND
03/03/2024 14:56:32.117 PST	DROF
03/03/2024 14:56:32.106 PST	DROF
03/03/2024 14:56:32.101 PST	DROF
03/03/2024 14:56:32.093 PST	DROF
03/03/2024 14:56:32.091 PST	DROF
03/03/2024 14:56:32.089 PST	DROF
03/03/2024 14:56:32.087 PST	DROF
03/03/2024 14:56:32.085 PST	DROF
03/03/2024 14:56:32.084 PST	DROF
03/03/2024 14:56:32.081 PST	DROF
03/03/2024 14:56:32.078 PST	DROF
03/03/2024 14:56:32.077 PST	DROF
03/03/2024 14:56:32.075 PST	DROF
03/03/2024 14:56:32.070 PST	DROF

Provenance Event

DETAILS

ATTRIBUTES

CONTENT

Input Claim

Container default

Section 3

Identifier 1709483775080-3

Offset 30548

Size 227 bytes

Download View

Output Claim

Container default

Section 3

Identifier 1709483775080-3

Offset 30548

Size 227 bytes

Download View

Replay

Connection Id 00b83ac0-018e-1000-3dfa-ffc2f7449971

OK

Step 3

Moved csv files to HDFS /data directory.

03

Into HDFS - /data

01-Kafka-PutHDFS

PutHDFS 1.25.0

org.apache.nifi-hdfs-hadoop-nar

0 (0 bytes)

5 min

Read/Write 0 bytes / 0 bytes

5 min

Out 0 (0 bytes)

5 min

Tasks/Time 0 / 00:00:00.000

5 min

Configure Processor | PutHDFS 1.25.0

Stopped

SETTINGS | SCHEDULING | PROPERTIES | RELATIONSHIPS | COMMENTS

Required field

Property	Value
Hadoop Configuration Resources	/Users/ag13/dsc650-infra/bellevue-bigdata/nifi/hado...
Kerberos Credentials Service	No value set
Kerberos User Service	No value set
Kerberos Principal	No value set
Kerberos Keytab	No value set
Kerberos Password	No value set
Kerberos Regoin Period	4 hours
Additional Classpath Resources	No value set
Directory	/data
Conflict Resolution Strategy	fail
Writing Strategy	Simple write
Block Size	No value set

NiFi Data Provenance

Displaying 29 of 29

Oldest event available: 03/01/2024 00:36:29 PST

Filter by component name

Date/Time	Type
03/03/2024 11:02:08.037 PST	DROF
03/03/2024 11:01:07.004 PST	DROF
03/03/2024 10:58:32.811 PST	DROF
03/03/2024 10:55:17.436 PST	DROF
03/03/2024 10:44:52.032 PST	DROF
03/03/2024 10:42:47.893 PST	DROF
03/03/2024 10:32:32.213 PST	DROF
03/03/2024 10:29:22.385 PST	DROF
03/03/2024 10:23:09.976 PST	DROF
03/03/2024 10:19:35.738 PST	DROF
03/03/2024 10:14:16.464 PST	DROF
03/03/2024 10:07:20.584 PST	DROF
03/03/2024 10:04:11.895 PST	DROF
03/03/2024 09:53:47.840 PST	DROF
03/03/2024 09:48:26.712 PST	DROF
03/03/2024 09:39:25.952 PST	DROF
03/03/2024 09:27:05.514 PST	DROF
03/03/2024 08:40:34.360 PST	DROF
03/03/2024 08:32:31.777 PST	DROF
03/02/2024 23:16:10.803 PST	DROF

Provenance Event

DETAILS | ATTRIBUTES | CONTENT

Input Claim

Container default

Section 3

Identifier 1709483775080-3

Offset 30321

Size 227 bytes

DOWNLOAD VIEW

Output Claim

Container default

Section 3

Identifier 1709483775080-3

Offset 30321

Size 227 bytes

DOWNLOAD VIEW

Replay

Connection Id 018c13d9-018e-1000-fbb9-7dafdb7fd74d

OK



#### 4. Data streaming to Kafka Producer

The screenshot displays the Apache NiFi web console. On the left, a flow diagram shows a process group named '04' containing a 'PublishKafka\_2\_6' processor. The processor's status is 'Stopped'. On the right, the 'Configure Processor' dialog for 'PublishKafka\_2\_6 1.25.0' is open. The 'PROPERTIES' tab is selected, showing a table of configuration properties. The 'Required field' section is visible at the top of the properties table.

Property	Value
Kafka Brokers	localhost:9092
Topic Name	users
Use Transactions	true
Transactional Id Prefix	No value set
Message Demarcator	No value set
Failure Strategy	Route to Failure
Delivery Guarantee	Guarantee Replicated Delivery
Attributes to Send as Headers (Regex)	No value set
Message Header Encoding	UTF-8
Security Protocol	PLAINTEXT
SASL Mechanism	GSSAPI
Kerberos Credentials Service	No value set

## NiFi Data Provenance

[illegible]

## Data streaming to Kafka Producer

The screenshot displays the Nifi console interface. On the left sidebar, there are sections for "Navigate" (with search, zoom, and view icons) and "Operate" (with a thumbs up icon). The main panel shows a flow diagram titled "Nifi -Kakfa Consumers" (note the typo in the original image). Below the flow diagram, a summary box for "Kafka Bids" (also a typo) provides details for "ConsumeKafka\_2\_6 1.25.0":

- In: 0 (0 bytes) / 5 min
- Read/Write: 0 bytes / 0 bytes / 5 min
- Out: 0 (0 bytes) / 5 min
- Tasks/Time: 0 / 00:00:00.000 / 5 min

The right-hand pane is titled "Configure Processor | ConsumeKafka\_2\_6 1.25.0". It features a red status indicator labeled "Stopped". Below this are tabs for "SETTINGS", "SCHEDULING", "PROPERTIES", "RELATIONSHIPS", and "COMMENTS". The "PROPERTIES" tab is active, displaying a table of properties:

Property	Value
Kafka Brokers	localhost:9092
Topic Name(s)	users
Topic Name Format	names
Group ID	01-nifi-consumer
Commit Offsets	true
Max Uncommitted Time	1 secs
Honor Transactions	true
Message Demarcator	No value set
Separate By Key	false
Security Protocol	PLAINTEXT
SASL Mechanism	GSSAPI
Kerberos Credentials Service	No value set

## NiFi Data Provenance

Displaying 148 of 148

Oldest event available: 03/01/2024 00:36:29 PST

events that match the specified qu

[illegible]

The diagram illustrates a data pipeline in NiFi. It starts with two input sources: 'CSV Files Category, Bids, Auctions' and 'NiFi Generator for test data Bidders, Auction List, Bids'. These feed into '01 Generate Kafka Producers' and '02 Generate Kafka Topic'. The data then flows through '01-Kafka-Producer', '01-Kafka-Consumer', and '01-Consumer-LogAttribute' to '01-Kafka-Producer' and '01-Kafka-Consumer'. The final output is 'LogAttribute'.

Based on the files in Hadoop – Hive queries are executed to find top bidders based on item category. Hive queries are also run based on zip code to segregate users from each state to get analytics of active users, revenue amount of each state.

## List of closed auctions from auctions table with user email addresses &amp; final bid price

```
hive> SELECT a.auction_id,a.item_description,a.category, a.final_bidder_internal_id,b.name,b.email,a.final_bid_price,a.bid_status FROM auctions
a JOIN bidders b ON (a.final_bidder_internal_id=b.user_internal_id) WHERE a.bid_status='closed';
Query ID = root_20240304022457_cbde13ce-85e3-4062-9765-f7a2baa64819
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1709518678473_0001)
```

	VERTICES	MODE	STATUS	TOTAL	COMPLETED	RUNNING	PENDING	FAILED	KILLED
Map 2 .....	container	SUCCEEDED	1	1	0	0	0	0	0
Map 1 .....	container	SUCCEEDED	1	1	0	0	0	0	0
VERTICES: 02/02 [=====]>>>] 100% ELAPSED TIME: 9.07 s									
OK									
10021	14KT White Gold 1.81 ctw Fancy Brown Diamond Ring	Building Materials	1016	ablotkamp	ablotkamp@1127Z.com	1157			
10044	Pearl and Diamond Necklace - 14KT Yellow Gold	Jewelry 1045	chueychuey	chueychuey@1026J.com	2652	closed			
10009	Rolux Datejust Watches 1037	chrisreed52@1059T.com	8175	closed					
10019	EarthQuake Benefit & Charity	1028	bartcee@1048J.com	7099	closed				
10041	Pearl and Diamond Necklace - 14KT White Gold	Gemstones	1027	barrettdonna	barrettdonna@1083X.com	7153	closed		
Time taken: 10.242 seconds, Fetched: 5 row(s)									
hive> █									

## Query 2:

Select bids that need to be emailed to a customer, from the bids table that are active with a minimum price that is above last price that a bidder has successfully closed the last. (Work in progress to write relevant sub queries etc.)

```
select a.auction_id,a.item_description,a.category, a.min_price, a.bid_status FROM auctions a join
(select min(final_bid_price) as max_price from auctions) b on b.max_price < a.min_price where
a.bid_status = 'active';
```

```
hadoop-hive-spark-hbase — docker-compose < docker compose exec master bash — 1...

hive> select a.auction_id,a.item_description,a.category, a.min_price, a.bid_status FROM auctions a join (select min(final_bid_p
ice) as max_price from auctions) b on b.max_price < a.min_price where a.bid_status = 'active';
Warning: Map Join MAPJOIN[16][bigTable=?] in task 'Reducer 3' is a cross product
2024-03-04 03:07:47,136 INFO [5a3b102d-a401-4695-812f-d350639e2836 main] reducesink.VectorReduceSinkEmptyKeyOperator: VectorRedu
ceSinkEmptyKeyOperator constructor vectorReduceSinkInfo org.apache.hadoop.hive.q1.plan.VectorReduceSinkInfo@9b84747
2024-03-04 03:07:47,138 INFO [5a3b102d-a401-4695-812f-d350639e2836 main] reducesink.VectorReduceSinkEmptyKeyOperator: VectorRedu
ceSinkEmptyKeyOperator constructor vectorReduceSinkInfo org.apache.hadoop.hive.q1.plan.VectorReduceSinkInfo@60580e42
Query ID = root_20240304030746_e438516d-9f7e-4e4d-9be4-2b0efefbce8b
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1709518678473_0008)

-----
VERTICES      MODE      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
-----
Map 2 ..... container  SUCCEEDED    1         1         0         0         0         0
Map 1 ..... container  SUCCEEDED    1         1         0         0         0         0
Reducer 3 ..... container  SUCCEEDED    1         1         0         0         0         0
-----
VERTICES: 03/03 [=====] 100% ELAPSED TIME: 9.67 s
-----
OK
10001 Tradition Idli Steamer Small Houseware 36 active
10050 Chinese Dinner Set Small Houseware 50 active
10048 Remodelled ADU Personal Property 35000 active
10047 Site in LakeTahoe Personal Property 250000 active
10043 Pearl and Diamond Necklace - 14KT Yellow Gold Jewelry 3420 active
10042 Pearl and Diamond Necklace - 14KT White Gold$1.00 Jewelry 9149 active
10037 Ikea Computer Desk Furniture 45 active
10036 Dinning Table for 8 Furniture 120 active
10035 Honda Two Wheeler Estate & Personal Property 5000 active
10034 EV Scooter Estate & Personal Property 450 active
10033 Samsung Sound Bar Electronics 5129 active
10032 Bose Audio Electronics 5696 active
10030 Google Chrome Book 2021 Computers & Electronics 258 active
10029 Macbook Pro Computers & Electronics 2494 active
10028 iMac 2020 Computers 6876 active
10027 Dell Inspiron Computers 2632 active
10023 Tesla S3 Cars & Vehicles 50000 active
10022 Tesla Cyber truck Cars & Vehicles 165000 active
10020 "30" Joits" Building Materials 4165 active
10018 SavePlanet Benefit & Charity 9975 active
10017 RedCross Shirts Benefit & Charity 856 active
10013 GE DishWasher Appliances 3803 active
10012 Samsung Microwave Appliances 4325 active
10008 Rolex Submariner Watches 5000 active
10007 Rolex Watches 3977 active
10005 DogMan Series Used Books & Magazines 75 active
10004 Magic Tree House Series Used Books & Magazines 200 active
10003 PingPong Table Sporting Goods 180 active
10002 Golfing Cart Sporting Goods 120 active
Time taken: 10.786 seconds, Fetched: 29 row(s)
hive>
```

## Conclusion

The project can be further enhanced by incorporating sending emails to the current bidders based on Hive queries calculation of which top 20 bidders performed better historically in a specific Item category.