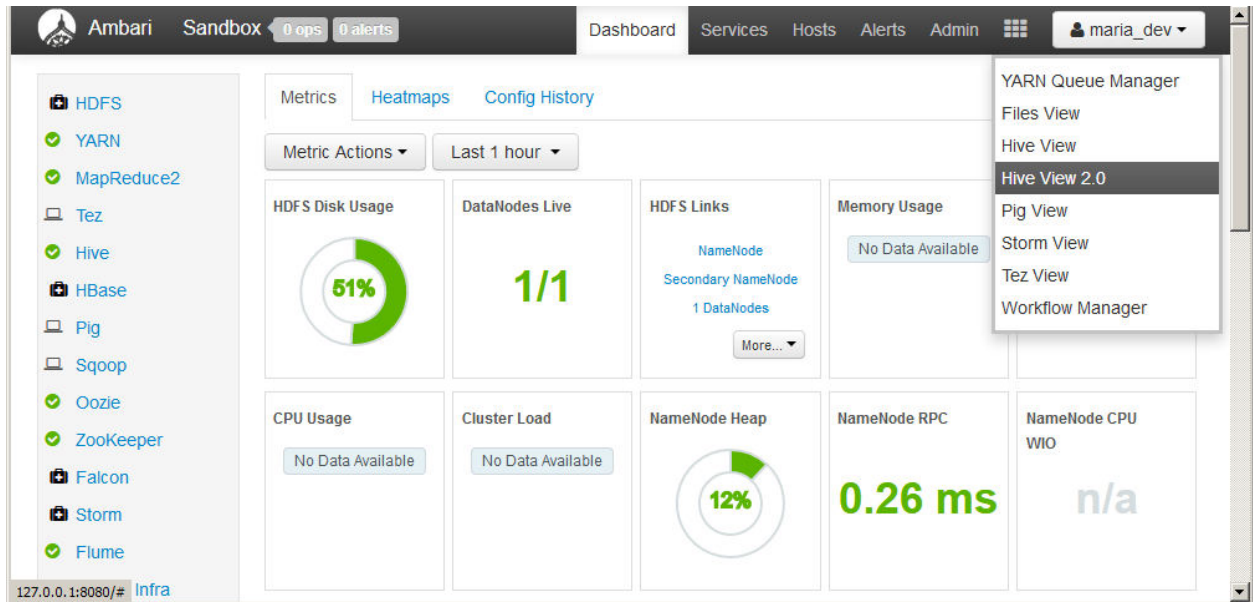




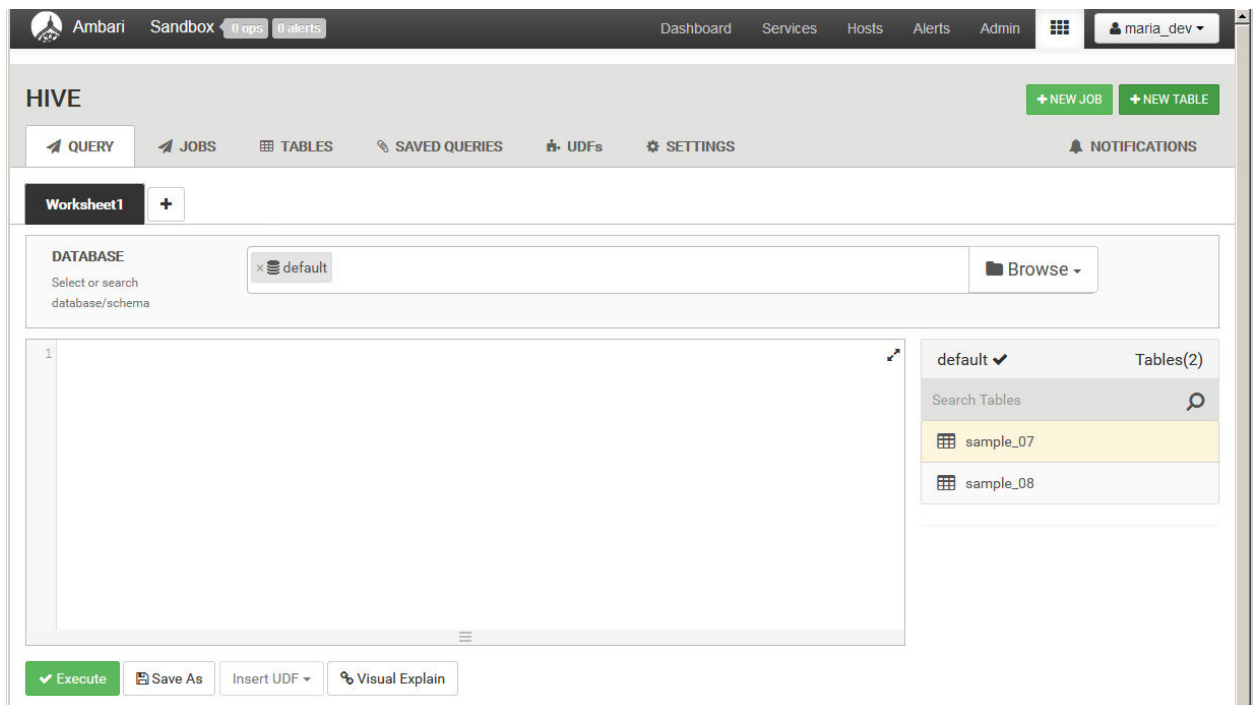
HiveQL

Hive – jest to platforma używana do tworzenia skryptów typu SQL do wykonywania operacji MapReduce.

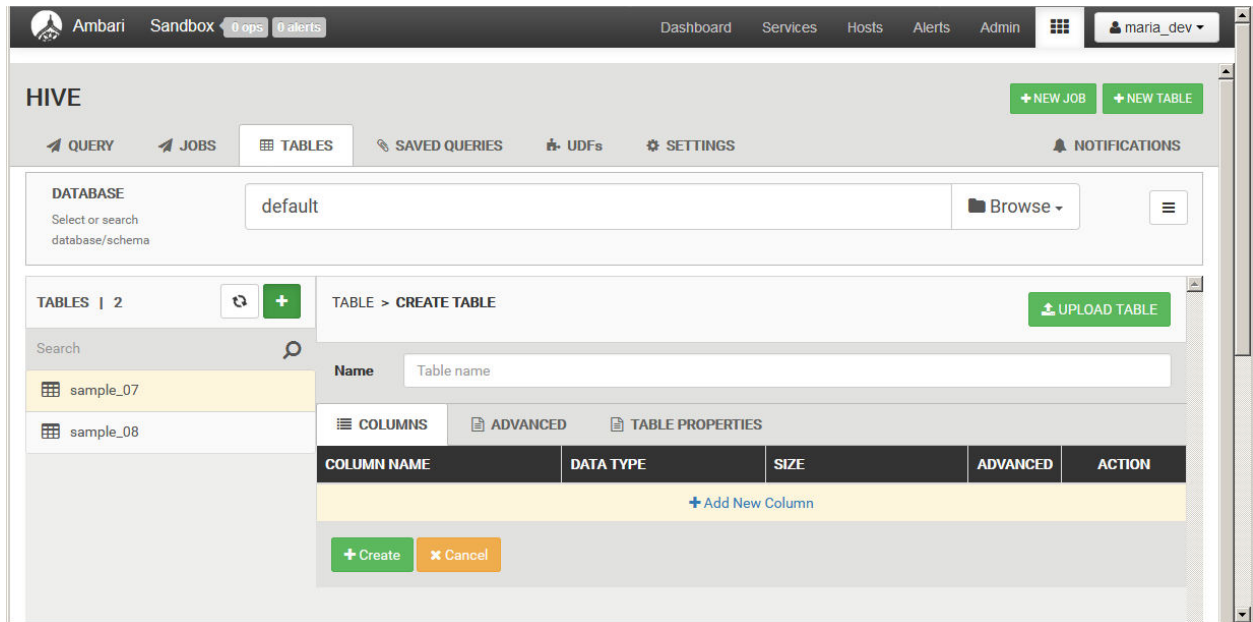
Otwórz 'Hive View 2.0' klikając w ikonę siatki w prawym górnym rogu:



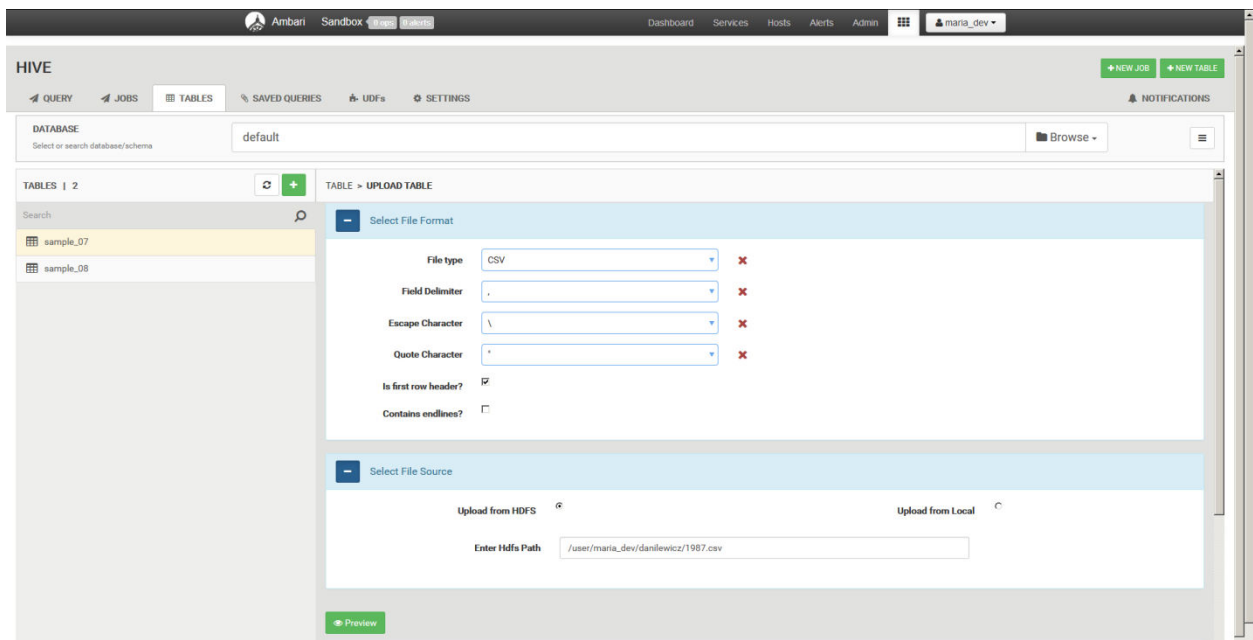
Kliknij '+ NEW TABLE', aby dodać nową tabelę:



Kliknij 'UPLOAD TABLE', aby załadować tabelę:



W pierwszej linii naszego pliku mamy nagłówki, więc zaznacz 'Is first row header? '. W sekcji 'Select File Source ' zaznacz 'Upload from HDFS ', a w 'Enter Hdfs Path ' wpisz ścieżkę do swojego pliku. U mnie jest to: /user/maria_dev/1987.csv



Kliknij 'Preview' i zamień nazwę tabeli z '1987' na 'loty'.

Wyższa Szkoła Bankowa

Gdańsk

Gdynia

Ambari

Sandbox

Create

Execute

Dashboard

Services

Hosts

Alerts

Admin

maria_dev

HIVE

QUERY

JOBS

TABLES

SAVED QUERIES

UDFs

SETTINGS

NEW JOB

NEW TABLE

NOTIFICATIONS

DATABASE

Select or search database/schema

default

Browse

Preview

Table Preview

| Year | Month | DayOfMonth | DayOfWeek | DepTime | CRSDepTime | ArrTime | CRSArrTime | UniqueCarrier | FlightNum | TailNum | ActualElapsedTime | CRSElapsedTime | AirTime | AirDelay | DepDel |
|------|-------|------------|-----------|---------|------------|---------|------------|---------------|-----------|---------|-------------------|----------------|---------|----------|--------|
| 1987 | 10 | 14 | 3 | 741 | 730 | 912 | 849 | PS | 1451 | NA | 91 | 79 | NA | 23 | 11 |
| 1987 | 10 | 15 | 4 | 729 | 730 | 903 | 849 | PS | 1451 | NA | 94 | 79 | NA | 14 | -1 |
| 1987 | 10 | 17 | 6 | 741 | 730 | 918 | 849 | PS | 1451 | NA | 97 | 79 | NA | 29 | 11 |
| 1987 | 10 | 18 | 7 | 729 | 730 | 847 | 849 | PS | 1451 | NA | 78 | 79 | NA | -2 | -1 |
| 1987 | 10 | 19 | 1 | 749 | 730 | 922 | 849 | PS | 1451 | NA | 93 | 79 | NA | 33 | 19 |
| 1987 | 10 | 21 | 3 | 728 | 730 | 848 | 849 | PS | 1451 | NA | 80 | 79 | NA | -1 | -2 |
| 1987 | 10 | 22 | 4 | 728 | 730 | 852 | 849 | PS | 1451 | NA | 84 | 79 | NA | 3 | -2 |
| 1987 | 10 | 23 | 5 | 731 | 730 | 902 | 849 | PS | 1451 | NA | 91 | 79 | NA | 13 | 1 |
| 1987 | 10 | 24 | 6 | 744 | 730 | 908 | 849 | PS | 1451 | NA | 84 | 79 | NA | 19 | 14 |
| 1987 | 10 | 25 | 7 | 729 | 730 | 851 | 849 | PS | 1451 | NA | 82 | 79 | NA | 2 | -1 |

Name: loty

COLUMNS

| COLUMN NAME | DATA TYPE | SIZE | ADVANCED | ACTION |
|-------------|-----------|------|----------|--------|
| Year | INT | | | Delete |

Potwierdź utworzenie tabeli klikając 'Create'. Proces ten może potrwać minute lub dwie.

Ambari

Sandbox

Create

Execute

Dashboard

Services

Hosts

Alerts

Admin

maria_dev

HIVE

QUERY

JOBS

TABLES

SAVED QUERIES

UDFs

SETTINGS

NEW JOB

NEW TABLE

NOTIFICATIONS

DATABASE

Select or search database/schema

default

Browse

| | | | | |
|-------------------|--------|--|--|--------|
| Origin | STRING | | | Delete |
| Dest | STRING | | | Delete |
| Distance | INT | | | Delete |
| TaxiIn | STRING | | | Delete |
| TaxiOut | STRING | | | Delete |
| Cancelled | INT | | | Delete |
| CancellationCode | STRING | | | Delete |
| Diverted | INT | | | Delete |
| CarrierDelay | STRING | | | Delete |
| WeatherDelay | STRING | | | Delete |
| NASDelay | STRING | | | Delete |
| SecurityDelay | STRING | | | Delete |
| LateAircraftDelay | STRING | | | Delete |

Add New Column

Create Cancel

Jeśli wszystko się udało, to wróćmy do zakładki 'QUERY'. Napišemy teraz naszą pierwszą kwerendę, która zliczy nam liczbę lotów w poszczególnych miesiącach. W oknie wpisz poniższą komendę i kliknij 'Execute'.

```
SELECT Month, COUNT(Month) AS Liczba_Lotow
FROM loty
GROUP BY Month;
```

The screenshot shows the Hive web interface. At the top, there's a navigation bar with 'Ambari', 'Sandbox', and 'alerts'. Below that, the 'HIVE' header is visible with buttons for '+ NEW JOB' and '+ NEW TABLE'. The main area is divided into tabs: 'QUERY', 'JOBS', 'TABLES', 'SAVED QUERIES', 'UDFs', and 'SETTINGS'. The 'QUERY' tab is active, showing a worksheet named 'Worksheet1'. The query editor contains the following SQL:

```
1 SELECT Month, COUNT(Month) AS Liczba_Lotow
2 FROM loty
3 GROUP BY Month;
```

Below the query editor, there are buttons for 'Execute', 'Save As', 'Insert UDF', and 'Visual Explain'. The 'RESULTS' tab is selected, showing a table with the following data:

| month | liczba_lotow |
|-------|--------------|
| 10 | 448620 |
| 11 | 422803 |
| 12 | 440403 |

Wynik kwerendy możemy zapisać do 'Widoku' (ang. 'View') dodając jedną linię na początku naszej kwerendy:

```
CREATE VIEW IF NOT EXISTS loty_w_miesiacu AS
SELECT Month, COUNT(Month) AS Liczba_Lotow
FROM loty
GROUP BY Month;
```

Widok jest wirtualną tabelą, którą usuwa się w następujący sposób:

```
DROP VIEW loty_w_miesiacu;
```

Aby usunąć tabelę wpisz:

```
DROP TABLE loty;
```