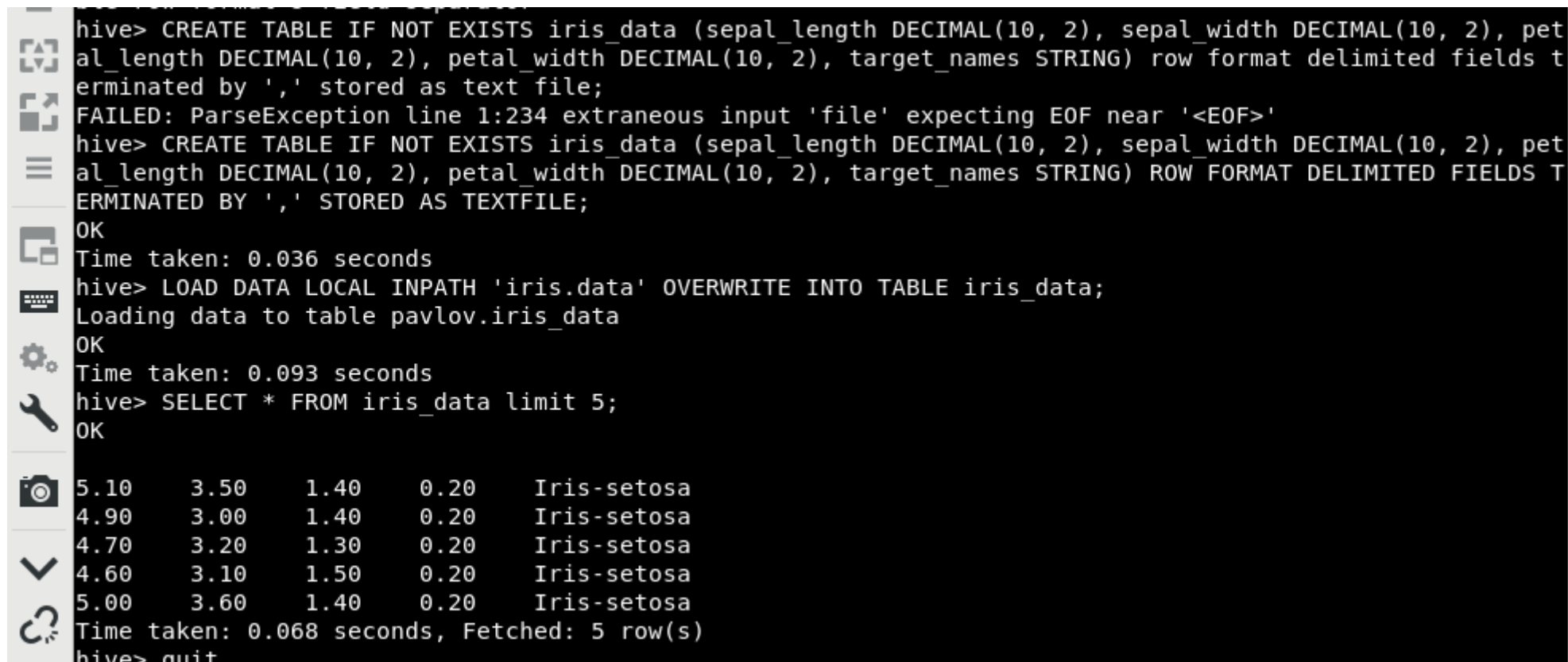


## Лабораторная работа №4

d) Создайте таблицу iris\_data

e) Загрузите данные из файла iris.data (ссылка на данные) в созданную таблицу iris\_data



```
hive> CREATE TABLE IF NOT EXISTS iris_data (sepal_length DECIMAL(10, 2), sepal_width DECIMAL(10, 2), petal_length DECIMAL(10, 2), petal_width DECIMAL(10, 2), target_names STRING) row format delimited fields terminated by ',' stored as text file;
FAILED: ParseException line 1:234 extraneous input 'file' expecting EOF near '<EOF>'
hive> CREATE TABLE IF NOT EXISTS iris_data (sepal_length DECIMAL(10, 2), sepal_width DECIMAL(10, 2), petal_length DECIMAL(10, 2), petal_width DECIMAL(10, 2), target_names STRING) ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' STORED AS TEXTFILE;
OK
Time taken: 0.036 seconds
hive> LOAD DATA LOCAL INPATH 'iris.data' OVERWRITE INTO TABLE iris_data;
Loading data to table pavlov.iris_data
OK
Time taken: 0.093 seconds
hive> SELECT * FROM iris_data limit 5;
OK
5.10      3.50      1.40      0.20      Iris-setosa
4.90      3.00      1.40      0.20      Iris-setosa
4.70      3.20      1.30      0.20      Iris-setosa
4.60      3.10      1.50      0.20      Iris-setosa
5.00      3.60      1.40      0.20      Iris-setosa
Time taken: 0.068 seconds, Fetched: 5 row(s)
hive> quit
```

f) Выполните следующие SQL выражения

- Вывести уникальные типы цветов ирисов на экран используя SQL запрос

```
hive> SELECT DISTINCT target_names FROM iris_data ORDER BY target_names;
```

```
Query ID = pavlov_20230801211727_7c09aled-aa31-4b13-88a1-676d729666f3
```

```
Total jobs = 1
```

```
Launching Job 1 out of 1
```

```
In order to change the average load for a reducer (in bytes):
```

```
  set hive.exec.reducers.bytes.per.reducer=<number>
```

```
In order to limit the maximum number of reducers:
```

```
  set hive.exec.reducers.max=<number>
```

```
In order to set a constant number of reducers:
```

```
  set mapreduce.job.reduces=<number>
```

```
Running with YARN Application = application_1690901753071_0011
```

```
Kill Command = /usr/lib/hadoop/bin/yarn application -kill application_1690901753071_0011
```

```
Hive on Spark Session Web UI URL: http://vm-dlake2-s-1.test.local:43342
```

```
Query Hive on Spark job[0] stages: [0, 1, 2]
```

```
Spark job[0] status = RUNNING
```

STAGES	ATTEMPT	STATUS	TOTAL	COMPLETED	RUNNING	PENDING	FAILED
Stage-0 .....	0	FINISHED	1	1	0	0	0
Stage-1 .....	0	FINISHED	1	1	0	0	0
Stage-2 .....	0	FINISHED	1	1	0	0	0

```
STAGES: 03/03 [=====>>] 100% ELAPSED TIME: 4.04 s
```

```
Spark job[0] finished successfully in 4.04 second(s)
```

```
OK
```

```
Iris-setosa
```

```
Iris-versicolor
```

```
Iris-virginica
```

```
Time taken: 17.474 seconds, Fetched: 3 row(s)
```

```
hive> █
```

- Найти среднее по первому атрибуту данных файла, используя SQL запрос

```
hive> SELECT AVG(sepal_length) AS avg_sepal_length FROM iris_data;
Query ID = pavlov_20230801211914_f439ff90-8779-4a4f-88c2-102e9233afe6
Total jobs = 1
Launching Job 1 out of 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
-----
      STAGES      ATTEMPT      STATUS      TOTAL      COMPLETED      RUNNING
-----
Stage-3 .....          0      FINISHED          1          1          0
Stage-4 .....          0      FINISHED          1          1          0
-----
STAGES: 02/02      [=====>>] 100% ELAPSED TIME: 1
-----
Spark job[1] finished successfully in 1.00 second(s)
OK
5.84333333333333333333333333333333
Time taken: 1.288 seconds, Fetched: 1 row(s)
hive>
```

- Найти min, max элемент по всем числовым столбцам в рамках группы цветов ирисов, используя SQL запрос

```
hive> SELECT target_names, MIN(sepal_length) AS min_sepal_length, MAX(sepal_length) AS max_sepal_length, MIN(sepal_width) AS
min_sepal_width, MAX(sepal_width) AS max_sepal_width, MIN(petal_length) AS min_petal_length, MAX(petal_length) AS max_petal
length, MIN(petal_width) AS min_petal_width, MAX(petal_width) AS max_petal_width FROM iris_data GROUP BY target_names ORDER
BY target_names;
Query ID = pavlov_20230801212155_1c548757-abe6-4012-9cfa-09c7fb26ce5c
Total jobs = 1
Launching Job 1 out of 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
-----

```

	STAGES	ATTEMPT	STATUS	TOTAL	COMPLETED	RUNNING	PENDING	FAILED
Stage-10 .....	0	FINISHED	1	1	0	0	0	
Stage-8 .....	0	FINISHED	1	1	0	0	0	
Stage-9 .....	0	FINISHED	2	2	0	0	0	

```
-----
STAGES: 03/03      [=====>>>] 100%  ELAPSED TIME: 1.00 s
-----

Spark job[3] finished successfully in 1.00 second(s)
OK
Iris-setosa      4.30    5.80    2.30    4.40    1.00    1.90    0.10    0.60
Iris-versicolor 4.90    7.00    2.00    3.40    3.00    5.10    1.00    1.80
Iris-virginica  4.90    7.90    2.20    3.80    4.50    6.90    1.40    2.50
Time taken: 1.115 seconds, Fetched: 3 row(s)
hive>
```