Hello Potential DreamMaker,

Thank you for taking the time to apply to the role of a Data Engineer

There are two questions, each designed to test a core competency

### Question 1

For this task, you will be fetching some metrics about weekend taxi rides in New York City between 2014 and 2016. You will be working on the New York City taxi trip dataset for 2009 to 2016 which is stored in a [Clickhouse](https://clickhouse.com/) database.

#### Task

Write an SQL query to fetch the following **monthly** metrics from the dataset for the interval between **1st January 2014** and **31st December 2016**:

* The average number of trips on Saturdays
* The average fare (**fare\_amount**) per trip on Saturdays
* The average duration per trip on Saturdays
* The average number of trips on Sundays
* The average fare (**fare\_amount**) per trip on Sundays
* The average duration per trip on Sundays

The output of the query should be in a format similar to this:

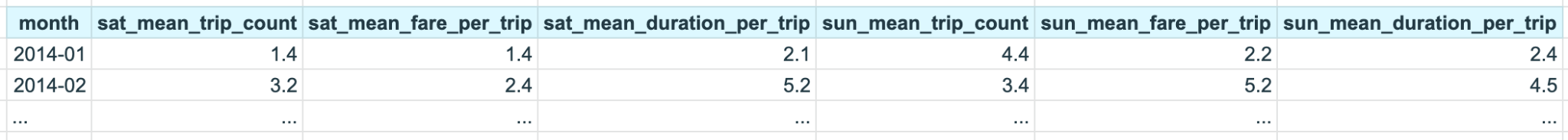


Fig 1. Query output format

#### How to Access the Database

The database can be accessed through a web browser by visiting the following link: <https://github.demo.trial.altinity.cloud:8443/play>. Use **demo** as your username and password (top left corner of the page).

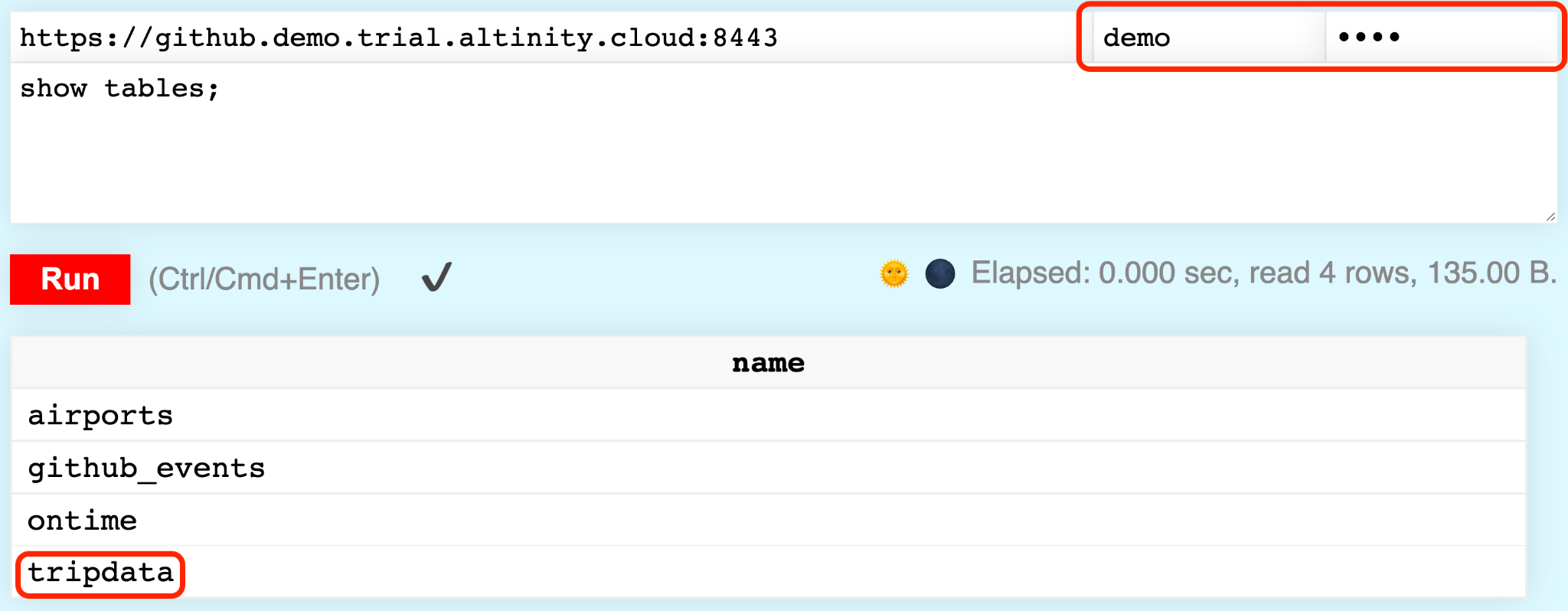


Fig. 2. Clickhouse web interface.

The dataset you will be working with is stored in the **tripdata** table. The column names are self explanatory, but note that we are interested in the **fare\_amount**, and not the other amount columns available in the table.

#### Submitting your Query

Save the query in a text file (.sql) and submit it along with the files from the subsequent question(s).

### Question 2

#### Task

This task follows from the previous one, but if you skipped it, you are free to mock the results as you see fit.

Your objective in this task is to build an Apache Airflow pipeline that gets the metrics you fetched from the previous question and writes the results into a table in a SQLite database.

For portability, it is recommended that you use an airflow instance with a SQLite database - the same one you will be writing to from your pipeline.

#### Deliverables

You are to hand in a zipped folder containing the following:

* DAG file(s)
* SQLite db file
* The query from Question 1
* Any other relevant files