# **NAME: NARENDRA GOLLA**

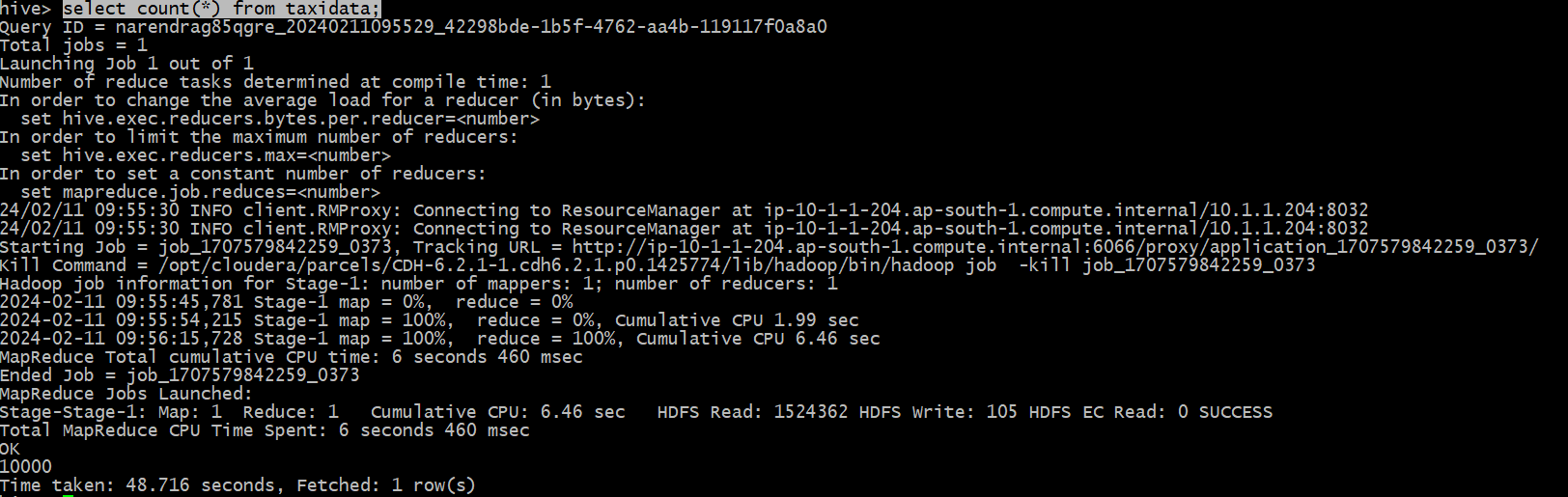
# **PROJECT: HIVE YELLOW TRIP**

## **Description**

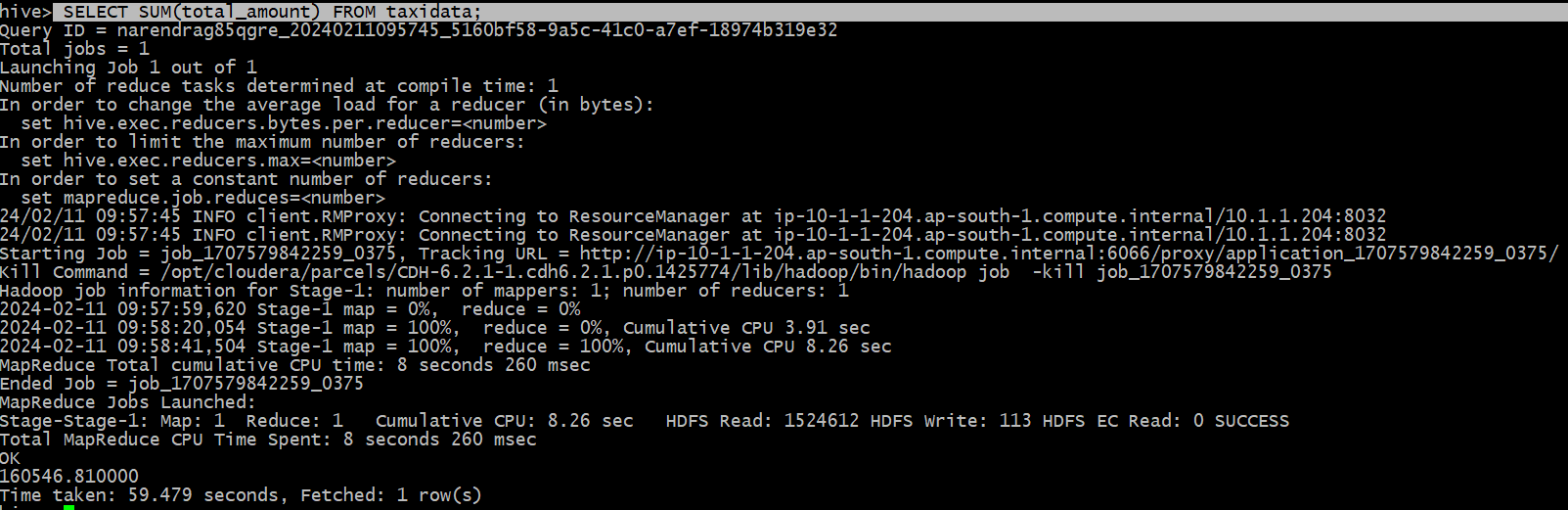
**Problem statement:**

In this case study, we are giving a real world example of how to use HIVE on top of the HADOOP for different exploratory data analysis. In here, we have a predefined dataset (2018\_Yellow\_Taxi\_Trip\_Data.csv) having more than 15 columns and more than 100000 records in it. \

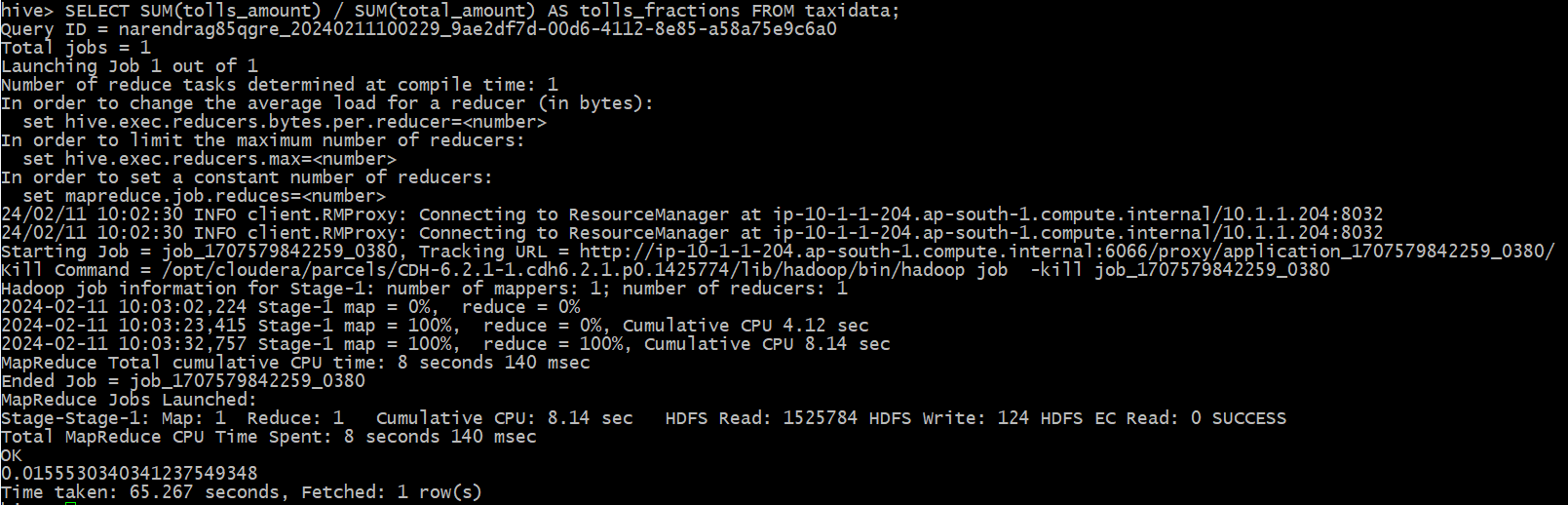
1. What is the total Number of trips ( equal to the number of rows)?



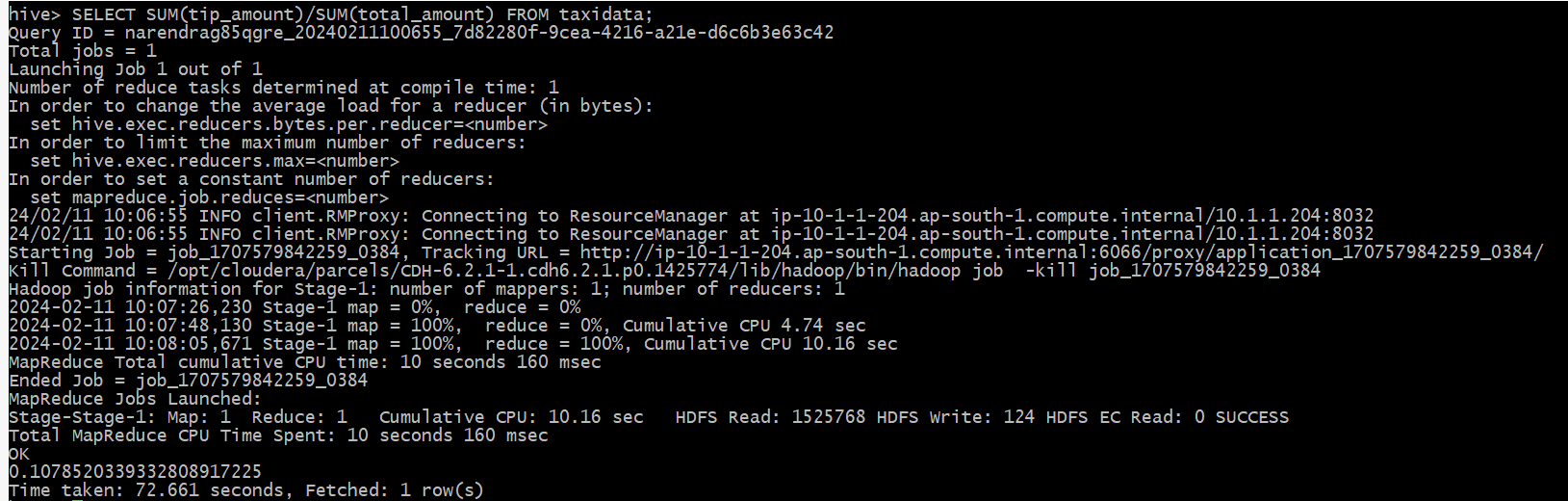
1. What is the total revenue generated by all the trips? The fare is stored in the column total\_amount.

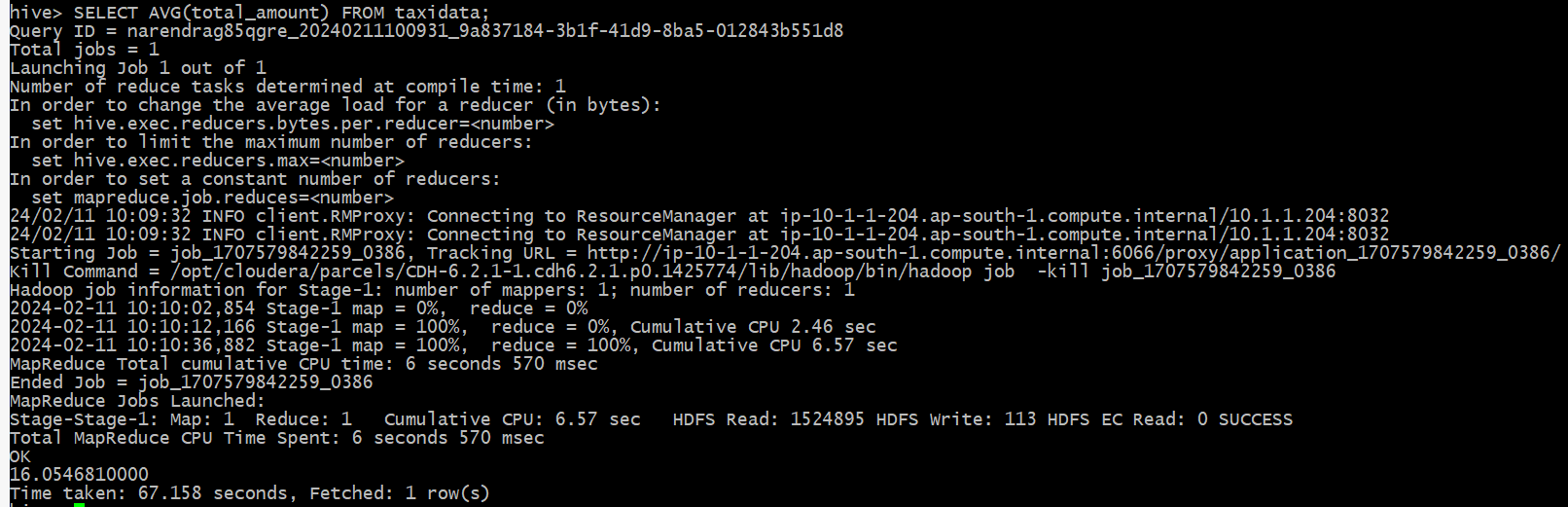


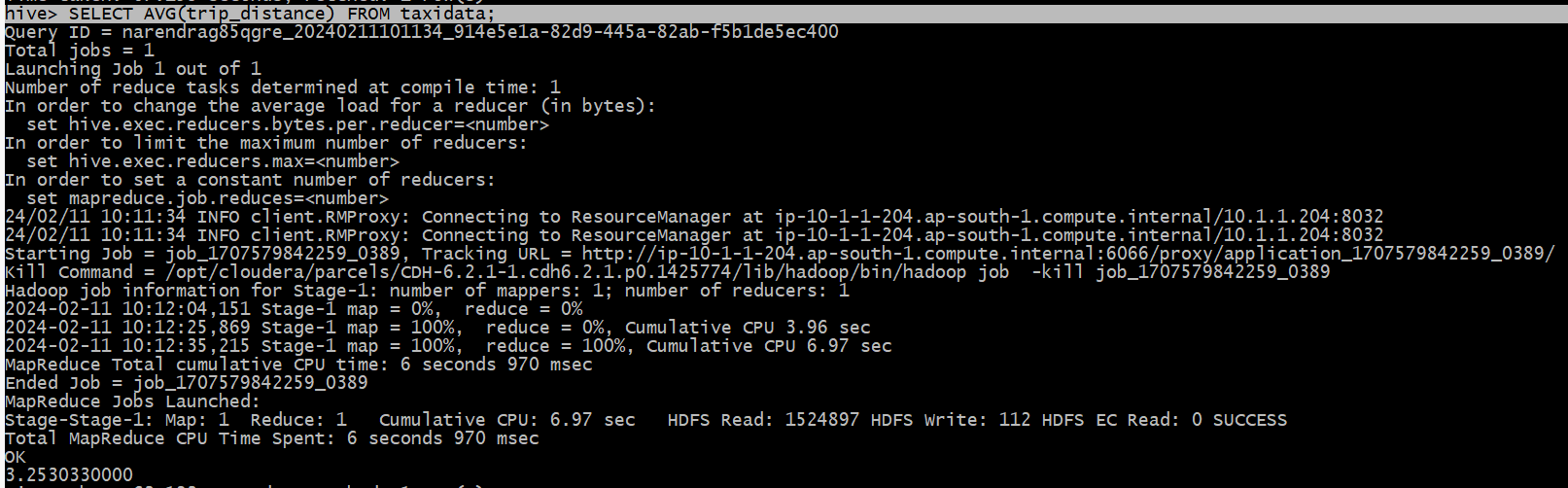
1. What fraction of the total is paid for tolls? The toll is stored in tolls\_amount.



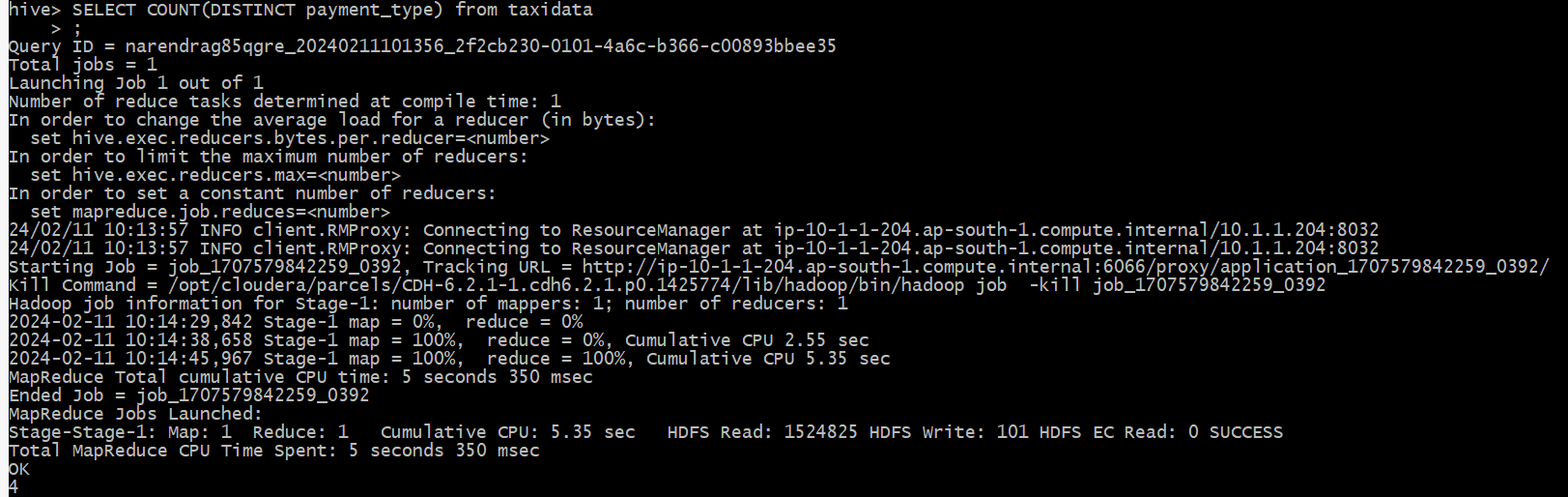
1. What fraction of it is driver tips? The tip is stored in tip\_amount.



1. What is the average trip amount? 
2. What is the average distance of the trips? Distance is stored in the column trip\_distance.

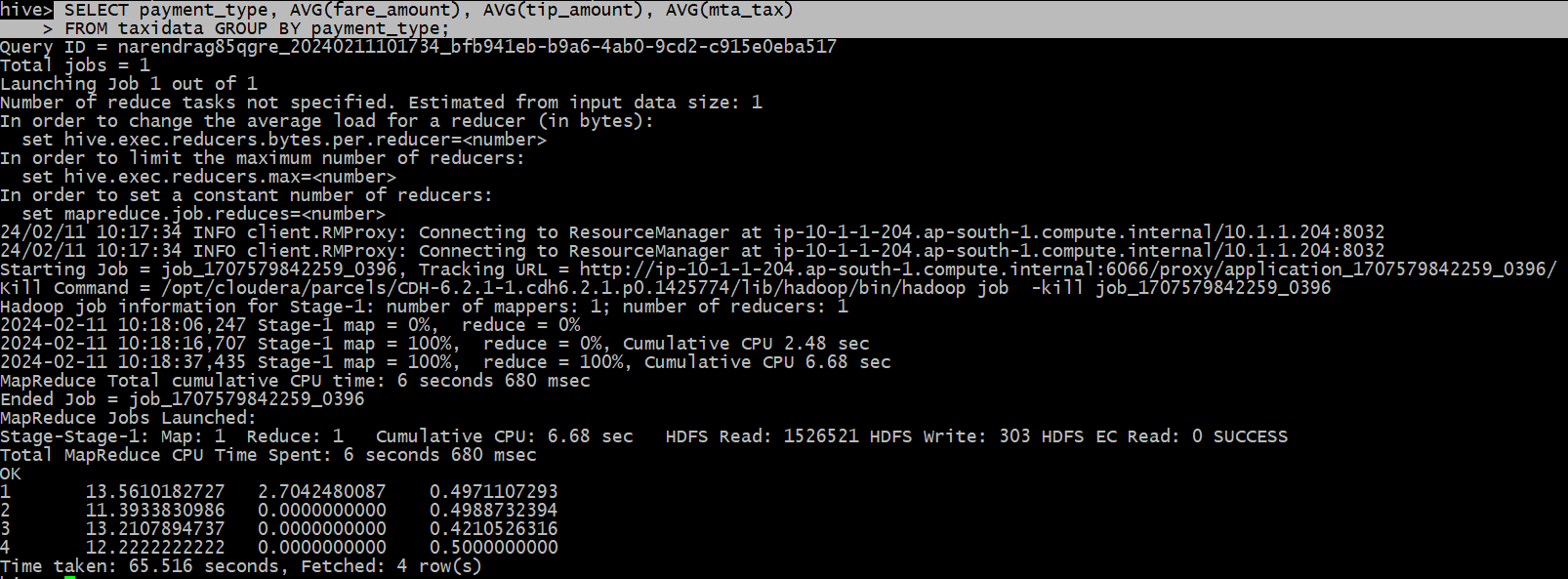


1. How many different payment types are used?



1. For each payment type, display the following details:

* Average fare generated
* Average tip
* Average tax – tax is stored in column mta\_tax



1. On an average which hour of the day generates the highest revenue?

