Assignment 4 - Shark

1. Launch the Shark shell.

```
shark
```

2. Create a table called *book* whose schema includes book's title, description, author's first name, last name, and cost.

```
create table
  book(title string, description string, firstname string, lastname string, cost int)
  row format delimited fields terminated by '\t';
```

3. List all the columns of the table *book*.

```
describe book;
```

4. Load the *book* table from the file *books* in the local filesystem. The *books* file has the following format:

Speed love	Long book about love	Brian	Dog	10
Long day	Story about Monday	Emily	Blue	20
Flying Car	Novel about airplanes	Phil	High	5
Short day	Novel about a day	Phil	Dog	30

```
load data local inpath 'books' into table book;
```

As an alternative solution, you can create the an *external* table. The external keyword lets you to create a table and provide a *location* so that Hive does not use a default location for this table. This would be useful if you already have data generated.

```
create external table
  exbook(title string, description string, firstname string, lastname string, cost int)
  row format delimited fields terminated by '\t'
  location '<file location, excluding the name of the file>';
```

5. Create a table called *novel*, containing those records from table *book* that have keyword "novel" in their description and cache it in memory.

```
create table novel TBLPROPERTIES('shark.cache'='MEMORY_ONLY')
  as select * from book where description like "%Novel%";
```

6. Print out the list of available tables.

```
show tables;
```

7. Count the number of records from the table *book*.

```
select count(*) from book;
```

8. Print out the total cost of the books with authors who have the same last name.

```
select lastname, sum(cost) from book group by lastname;
```

9. Count the number of distinct last names.

```
select count(distinct lastname) from book;
```

10. Drop the tables.

```
drop table book;
drop table novel;
```