Here's how I created a Hive table, imported data into it, and ran a simple search.

I used PuTTY Secure Copy to transfer a csv file to my CentOS VM:

C:\Program Files (x86)\PuTTY>pscp D:\state\_capitals.csv hadoop@192.168.1.123:/home/hadoop  
hadoop@192.168.1.123's password:  
state\_capitals.csv        | 1 kB |   1.5 kB/s | ETA: 00:00:00 | 100%

Next, I copied it into my Hadoop filesystem. If

[hadoop@localhost ~]$ hadoop fs -ls

returns an error like file or directory not found, you may have to re-create the /user/hadoop directory. Otherwise,

[hadoop@localhost ~]$ hadoop fs -mkdir states  
[hadoop@localhost ~]$ hadoop fs -copyFromLocal /home/hadoop/state\_capitals.csv states

It gets a little complicated, but bear with me. I created a temporary table in Hive that stores all the data in one field. You've seen this in Excel, where the data goes into column a and you have to parse it.

hive> CREATE TABLE temp\_capitals(data STRING);  
OK  
Time taken: 0.219 seconds

hive> LOAD DATA LOCAL INPATH 'state\_capitals.csv' INTO TABLE temp\_capitals;  
Loading data to table default.temp\_capitals  
Table default.temp\_capitals stats: [numFiles=1, totalSize=1521]  
OK  
Time taken: 0.232 seconds

Then I created the final table.

hive> CREATE TABLE state\_capitals(state\_name STRING, capital STRING, population INT);  
OK  
Time taken: 0.082 seconds

I extracted the data from my temporary table using wildcards, and inserted it into my table.

hive> INSERT OVERWRITE TABLE state\_capitals  
    > SELECT  
    > regexp\_extract(data, '^(?:([^,]\*)\,?){1}', 1) state\_name,  
    > regexp\_extract(data, '^(?:([^,]\*)\,?){2}', 1) capital,  
    > regexp\_extract(data, '^(?:([^,]\*)\,?){3}', 1) population  
    > FROM temp\_capitals;

Finally, I ran a conditional select statement to query the capital of Colorado

hive> SELECT  
    > capital  
    > FROM state\_capitals  
    > WHERE state\_name = 'Colorado';  
OK  
Denver  
Time taken: 0.101 seconds, Fetched: 1 row(s)  
hive>

Instead of the intermediate step of pulling the data into one field and then parsing from within hive, I found another way to just pull it in directly from the csv file you provided.  Here is the statement I used

hive> CREATE TABLE state\_capitals(state\_name STRING, capital STRING, population INT)row format delimited fields terminated by ',' stored as textfile;