# Assignment 2. HBase.

Provide screenshots of all steps. Add a brief description AS NEEDED.

1. If you are out of campus, establish VPN connection to FDU network. To do so, follow instructions as explained in Assignment 1.
   - If you are on campus, you don't need to establish a VPN connection.
2. For Windows users: download the Putty.exe SSH client from http://www.chiark.greenend.org.uk/~sgtatham/putty/download.html
   - Mac users may use the built-in shell. Note for Mac users: if any time during this or subsequent assignments you experience a problem, please use a Windows-based computer at any lab located at FDU Becton Hall.
3. Start Putty.exe and connect to the cluster name node by entering the following IP address 132.238.52.158.
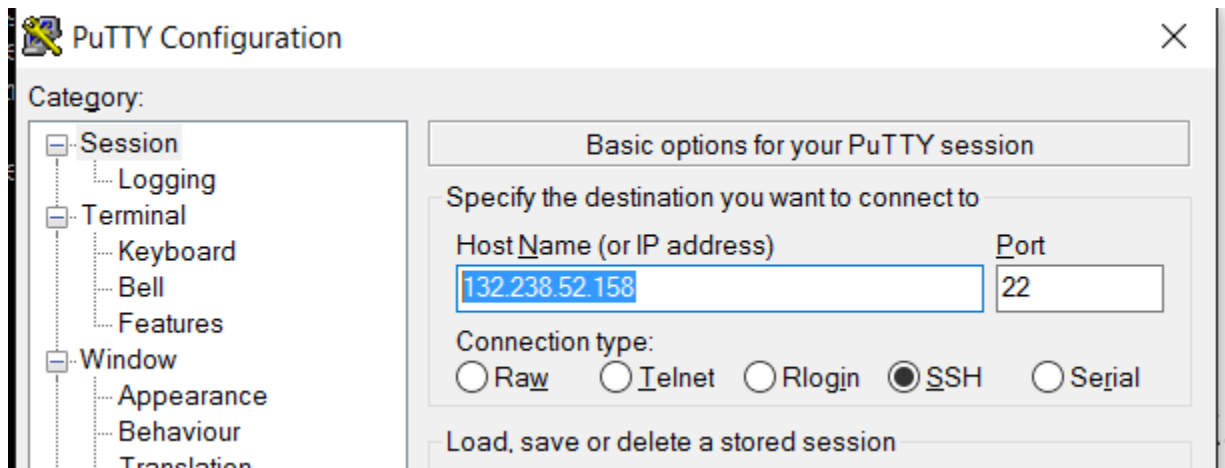
4. Click the [Open] button to establish a connection.
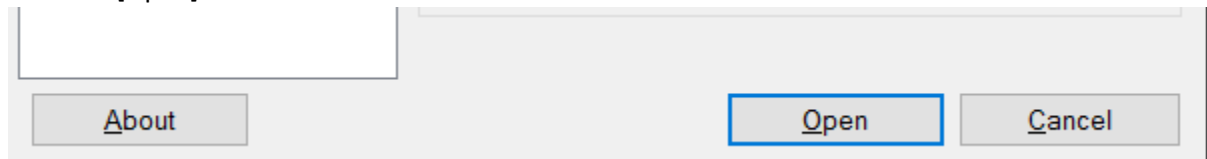
5. Login to the Apache Hadoop cluster using your ID and password.
6. Start HBase shell by executing the following command: **hbase shell** (you need to type hbase shell and press the [Enter] button.
7. Execute **status** command. It will show the status of HBase.
8. Check HBase version by running **version** command.
9. Get help on table-related commands by running **table_help** command.
10. Check your user information by running **whoami** command.
11. Create a table with a single column family with a single column family.
    - In the command below, substitute stXX with your Hadoop username:
    create 'cars_stXX', 'family_stXX'
12. Insert 5 rows in the table using the commands below. Make sure to substitute stXX, model_value, year_value, and color_value with your own values:

put 'cars_stXX', 'row1', 'family_stXX:model', 'model_value'
put 'cars_stXX', 'row1', 'family_stXX:year', 'year_value'

put 'cars_stXX', 'row1', 'family_stXX:color', 'color_value'
put 'cars_stXX', 'row2', 'family_stXX:model', 'model_value'
put 'cars_stXX', 'row2', 'family_stXX:year', 'year_value'
put 'cars_stXX', 'row2', 'family_stXX:color', 'color_value'
put 'cars_stXX', 'row3', 'family_stXX:model', 'model_value'
put 'cars_stXX', 'row3', 'family_stXX:year', 'year_value'
put 'cars_stXX', 'row3', 'family_stXX:color', 'color_value'
put 'cars_stXX', 'row4', 'family_stXX:model', 'model_value'
put 'cars_stXX', 'row4', 'family_stXX:year', 'year_value'
put 'cars_stXX', 'row4', 'family_stXX:color', 'color_value'
put 'cars_stXX', 'row5', 'family_stXX:model', 'model_value'
put 'cars_stXX', 'row5', 'family_stXX:year', 'year_value'
put 'cars_stXX', 'row5', 'family_stXX:color', 'color_value'

13. List the contents of the table (again, you need to substitute stXX with your Hadoop username):

scan 'cars_stXX'

14. List the setting of the table by running this command:

describe 'cars_stXX'

15. Display the fifth row of the table by running this command:

get 'cars_stXX', 'row5'

16. Delete cell value in row1 for column color:

delete 'cars_stXX', 'row1', 'family_stXX:color'

17. Using example above, delete cell value in row2 column year.
18. Updating a value. This is done through inserting a new value to a cell using the ***put*** command. Your task. update a value in row5 column color by doing this:
put 'cars_stXX', 'row5', 'family_stXX:color', 'newcolor'
19. Let's set HBase to keep 7 versions of data. To do this, we need to alter the table:

alter 'cars_stXX', NAME => 'cars_stXX', VERSIONS => 7

20. Now, update a value in row5 column color two times:

put 'cars_stXX', 'row5', 'family_stXX:color', 'blue'
put 'cars_stXX', 'row5', 'family_stXX:color', 'green'
21. Perform scan command again to list the contents of the table. Notice that only the latest version of the data is listed.
22. Now, let's see different versions of data with corresponding timestamps by running this:
scan 'cars_stXX', {RAW => true, VERSIONS => 5}
23. Exit HBase shell by running this command:
quit