Julie,

 a couple of things

  1.  in the loop that generates data,  it should only write the data to the file.

  2.  the table is not created until after the loop completes.  loadTable function does the job of reading the test data from the file and loading the table.

  3.  after loadTable,  then you can do getRowByKey to retrieve by key,  or sendToAll to do query on a non-key column.

  4.  also remember that the worker nodes have to started and running before you run mpclient

See additional comments in the code below.

David

|  |  |
| --- | --- |
|  | from cst363ona.src import mpclient import random |
|  |  |
|  | # create student table on all nodes |
|  | mpclient.readConfig() |
|  | f = open("student.data", "w") |
|  |  |
|  | # generate 100 rows of test data into a file and load the database |
|  | for i in range(1, 101): |
|  | name = 'Student' + str(i) |
|  | # Randomly generate student id |
|  | studentid=i |
|  | ''' |
|  | for a in range(1,400): |
|  | studentid=a+4 |
|  | ''' |
|  | #Major should be random selection of ‘Biology’, ‘Business’, ‘CS’, ‘Statistics’ |
|  | if i % 2: |
|  | major='Biology' |
|  | elif i % 3: |
|  | major='Statistics' |
|  | elif i % 5: |
|  | major='CS' |
|  | else: |
|  | major='Business' |
|  | #gpa is a random value from 2.5 to 4.0 |
|  | gpa=round(random.uniform(2.5, 4.0), 2) |
|  | #Table: student (id int primary key, name char(20), major char(10), double gpa) |
|  | line = str(studentid)+', "'+name+'", '+'"'+major+'", '+str(gpa)+'\n' |
|  |  |
|  | # Verify that the cluster is working by doing the queries in mpclient1.py |
|  | # ''' |
|  | **# c = mpclient.Coordinator() Don't do this during the loop to create test data.** |
|  | # select \* from student where id=2 -- 2 is an id that exists |
|  | print("test1") |
|  | # c.getRowByKey("select \* from student where studentid=" + str(studentid), '2') |
|  | if str(studentid) == '2': |
|  | print("two is present") |
|  | print("test2") |
|  | # select \* from student where id=yyyy -- yyyy is an id that does not exist |
|  | # c.getRowByKey("select \* from student where studentid=" + str(studentid), '150') |
|  | if str(studentid) == '150': |
|  | print("150 is not present") |
|  | print("test3") |
|  | # select \* from student where major = ‘Biology’ |
|  | # c.getRowByKey("select \* from student where major=" + str(major), 'Biology') |
|  | if major=='Biology': |
|  | print("we have bio majors") |
|  | print("test4") |
|  | # select \* from student where gpa >= 3.2 |
|  | # c.getRowByKey("select \* from student where gpa>=" + str(gpa), 3.20) |
|  | if gpa>=3.20: |
|  | print("we have honor roll") |
|  | #''' |
|  | f.write(line) |
|  | f.close() |
|  |  |
|  | c = mpclient.Coordinator() |
|  | c.sendToAll("drop table if exists student") |
|  | c.sendToAll("create table student (studentid int primary key, name char(20), major char(10),gpa double)") |
|  | c.loadTable("student","student.data") # now that table has been created and loaded, you can test getRowByKey studentid = 2 c.getRowByKey('select \* from student where studentid=' + str(studentid), studentid) # Biology is not the key. so query has to be sent to all nodes and results collected together c.sendToAll('select \* from student where major="Biology" ') c.sendToAll('select \* from student where gpa>=3.20') |
|  | c.close() |