Assignment 4 Integrate Cassandra

Hailun Zhu ID[hailunz]

1. Parameters setting:
   Number of the trees : N = 5;
   Number of the features: featureNum = 3, total 6 features. I choose ceiling $\sqrt{6}$ = 3.
2. Tables
   a) train

lineid[int, primary key]	f0[int]	f1[int]	f2[int]	f3[int]	f4[int]	f5[int]	label[int]

lineid: current line id, unique;
f0~ f5: features;
label: label.

b) test

lineid[int, primary key]	f0[int]	f1[int]	f2[int]	f3[int]	f4[int]	f5[int]	label[int]

lineid: current line id, unique;
f0~ f5: features;
label: label.

c) result (accuracy)

timestamp [timestamp] primary key	forest [double]	tree0 [double]	tree1 [double]	tree2 [double]	tree3 [double]	tree4 [double]

timestamp: indicates the current time of the test. Unique.
forest: accuracy of the whole forest.
tree0 ~ tree4: accuracies of the trees individually.

d) tree (tree object)

uuid [uuid] primary key	timestamp [timestamp]	length [int]	object [blob]

Timestamp: indicates the current time of the test.
I created an index on timestamp in order to select all the trees created in this test run.
I convert the TreeNode object to a bytebuffer and store it in the database. Length is the length of the bytebuffer.