Accuracy and performance comparison between HyperLogLog and FM Sketch

1. Physical Setup: my macbook pro, Intel core i7 2.3GHz, 16GB memory
2. HLL setup p=14, use parse register first and then dense registers
3. FM setup is bitvector=16, i.e., less than 20% error theoriatically.
4. x-axis means the true value of NDV.
5. y-axis means error or count time. Error means relative error, i.e., (TRUTH-COUNT)/TRUTH
6. NDV=0.1NUM means that assume that we have candidate number 1-10, we generate 10 numbers, however, 10 number will only have a NDV=0.1\*10=1, i.e., all the values are the same
7. Conclusions (1) HLL’s error is much less than that of FM in almost all the cases. When NDV=NUM, HLL’s error may be more than that of FM, but it is always less than 10% and it is stable. (2) HLL’s running time is always less than that of FM in all the cases.











