

What are the best k hashes and m bits values to store one million n keys (E.g. e52f43cd2c23bb2e6296153748382764) suppose we use the same MD5 hash key from [pickle\\_hash.py](#) and explain why?

For K, we have equation to calculate optimal number of hash functions  
 $k = (m/n) \ln(2)$

We using the same MD5 hash key which means false positive rate  $fp\_prob = 0.05$ .

$$fp\_ptob = (1 - e^{-kn/m})^k$$

$$n = 1,000,000$$

$$\text{After simplify, } K = -\ln(fp\_prob) / \ln(2) = 4$$

$$m = 5,771,172$$