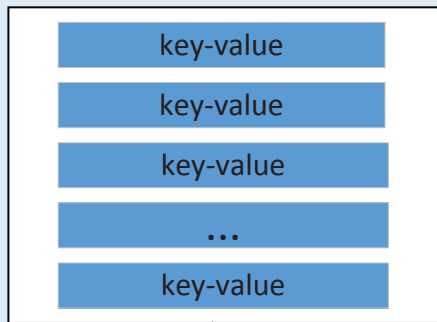


Source file



Writing

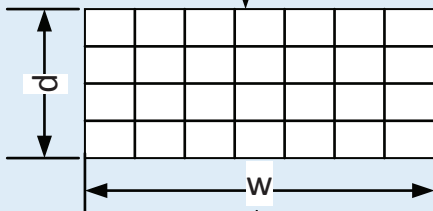
Output file

Big groups

Small groups

Counting

CM sketch



getRoughSize(key)

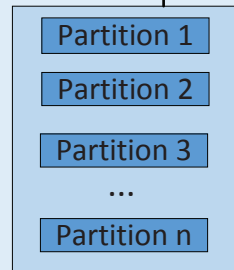
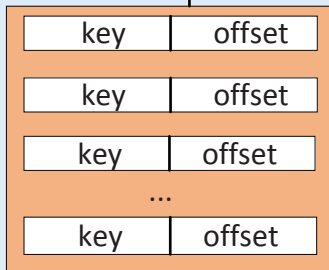
roughSize > threshold

Yes

No

Filling

Hashing



Indexing

Partition