



## Sheet 3: Partitioning

1. Mention one advantage and one disadvantage for each of the following techniques for rebalancing partitions,
  - a. Hash mode N
  - b. Fixed number of partitions
  - c. Dynamic Partitions with respect to data size
  - d. Dynamic Partitions with respect to nodes count
  - e. Consistent Hashing without virtual nodes
2. Given a cluster with N servers that uses Hash modulo N for rebalancing partitions. The cluster uses a perfectly uniform hash function .
  - a. For  $N = 4$ , what is the exact fraction of the objects that have to be reassigned when N changes by 1 (a server removed or added)
  - b. Retry the problem with  $N = 5$
  - c. Retry the problem with  $N = 6$
  - d. Can you drive a general formula for calculating this
  - e. Now assume the cluster will use Consistent Hashing with virtual nodes. Please repeat questions (a),(b),(c) and (d).
3. A nice property of consistent hashing is that it is easy to accommodate servers with more capacity. Say, for example, that a server has twice the amount of space than the others. How can we make this server receive twice as many objects than the others, if we are using consistent hashing?