Split index into RangePartitions based on load
Split at PartitionKey boundaries
PartitionMap tracks Index RangePartition assignment to partition servers
Front-End caches the PartitionMap to route user requests
Each part of the index is assigned to only one Partition Server at a time