* Join optimization :
  + Big table – small table : (relatively low cardinality) broadcast join
  + Big table – big table : (high cardinality) sort and bucket the data ahead of time then do SMB JOIN (sorted bucket merge join)
* Not to do shuffle
* Managing retention and storage