Advanced PushDown Optimizations

INTRODUCTION

Optimizes the table data, depending upon table size.

DESCRIPTION

Apache CarbonData pushes as much of query processing as possible close to the data to minimize the amount of data being read, processed, converted and transmitted/shuffled. Using projections and filters it reads only the required columns from the store and also reads only the rows that match the filter conditions provided in the query.

- In a join query if the table is greater than 10 MB then it will be pushed down which means CarbonData will keep that table with itself to process. There are two types of Push Down operations, that is Left Push Down and Right Push Down.
- Left Push Down: If the table on the left side in the join query is greater than the table on the right side and simultaneously greater than 10 MB then it will be a Left Push Down.
- Right Push Down: If the table on the right side of the join query is greater than the table on the left side and simultaneously greater than 10 MB then it will be a Right Push Down.

***Bucketing**: It is a technique that is used for uniform distribution of data across files in CarbonData. It enhances the performance of join queries. While loading the data, records are placed into buckets based on the hashing algorithm(s). During the execution of join queries the records can be fetched from buckets without the need of shuffling. This feature is used to distribute/organize the table/partition data into multiple files placing similar records in the same file.