Auto Compaction Design

Background:

Openlookeng not support auto compaction of tables. This becomes the limitation of openlookeng Since other processing engines like spark provide the same.

Objective：

System automatically initiates triggers to connecter to scans through all Db and tables and selects the for minor compaction tables. Gives the selected list of table to Engine task that do compaction.

Configuration

In config.properties below configuration need to be added

Name	Description	Default Value
auto-compaction.enabled	Enable(true) or disable(false) the auto compaction scan	false
auto-compaction.scan.interval	scan interval	15 sec
auto-compaction.scan.threads	scan thread number	5 threads

High Level Design

1) When coordinator module starts it will create one background thread, that thread will check auto-compaction is enabled or not.
   When it is enabled it will create threads depend up on auto-compaction.scan.threads and threads will start scanning with the interval auto-compaction.scan.interval.
2) For every scan interval it will scan all catalogs and assign to one of the thread pool.
3) These threads will used rest interface（v1/compaction/{catalogName}）to interact catalog module. When the node receive the scan request. It will run the compaction scan of the specified catalog. When the scan finish, return the scan result to scan thread.
4) When the scan thread receive the scan result. If will call the API in disptchManager to create a vacuum query and wait the query finished.

interaction auto-compaction-Engine

AutoCompactionScanner | startCompactionScan | CarbondataMetadata | HiveMetadata | DispatchManager

1 : run

if autocompaction is disabled return

1) Depending up on configuration create tasks. 2) schedule it based up on configuration

2 : scan

1) get all catalogs 2) assign to one of the thread

3 : startCompactionScan
Rest base Scan request /v1/compaction/{catalogName}

4 : startCompactionScan

5 : List String DbName.TableName

6 : startCompactionScan

7 : List String DbName.TableName

8 : List String DbName.TableName

9 : startCompaction

10 : createQuery

11 : getDispatchInfo

Keep waiting till query is finished