Carbondata Auto Cleanup

Background:

Openlookeng integration with Carbondata not supporting auto cleanup of vacuum/compaction tables. This becomes the limitation of openlookeng since other processing engine like spark provides the same.

Objective:

After vacuum/compaction operation is completed on carbon tables, there will be unused base/ stale files which are left in HDFS. So Carbondata Auto cleanup is used to cleanup those files automatically.

It is will clean up the COMPACTED files which crossed more than 60 minutes.

Auto cleanup will be triggered in case of insert, update, delete & vacuum.

After minor compaction

store/default/st2/Fact .store/default/st2/Fact/Part0 on.store/default/st2/Fact/Part0/Segment_0 on.store/default/st2/Fact/Part0/Segment 0/0 1592469655636.carbonindexmerge pn.store/default/st2/Fact/Part0/Segment 0/part-0-0 batchno0-0-0-1592469643544.carbondata on.store/default/st2/Fact/Part0/Segment_0.1 on.store/default/st2/Fact/Part0/Segment_0.1/0.1_1592477541574.carbonindexmerge on.store/default/st2/Fact/Part0/Segment_0.1/part-0-0_batchno0-0-0.1-1592477533846.carbondata on.store/default/st2/Fact/Part0/Segment_0.1/part-0-1_batchno0-0-0.1-1592477533846.carbondata on.store/default/st2/Fact/Part0/Segment_1 on.store/default/st2/Fact/Part0/Segment_1/1_1592470206156.carbonindexmerge on.store/default/st2/Fact/Part0/Segment_1/part-0-0_batchno0-0-1-1592470190145.carbondata on.store/default/st2/Fact/Part0/Segment 2 on.store/default/st2/Fact/Part0/Segment_2/2_1592470278245.carbonindexmerge on.store/default/st2/Fact/Part0/Segment_2/part-0-0_batchno0-0-2-1592470270215.carbondata on.store/default/st2/Fact/Part0/Segment_2/part-0-1_batchno0-0-2-1592470270215.carbondata on.store/default/st2/Fact/Part0/Segment_3 on.store/default/st2/Fact/Part0/Segment_3/3_1592470301948.carbonindexmerge on.store/default/st2/Fact/Part0/Segment_3/part-0-0_batchno0-0-3-1592470293826.carbondata on.store/default/st2/Fact/Part0/Segment_4 on.store/default/st2/Fact/Part0/Segment 4/4 1592470314996.carbonindexmerge on.store/default/st2/Fact/Part0/Segment_4/part-0-0_batchno0-0-4-1592470308901.carbondata on.store/default/st2/LockFiles on.store/default/st2/LockFiles/Segment_0.lock on.store/default/st2/LockFiles/Segment_1.lock on.store/default/st2/LockFiles/Segment_2.lock on.store/default/st2/LockFiles/Segment_3.lock on.store/default/st2/LockFiles/Segment_4.lock on.store/default/st2/LockFiles/compaction.lock on.store/default/st2/LockFiles/meta.lock on.store/default/st2/LockFiles/tablestatus.lock on.store/default/st2/LockFiles/update.lock on.store/default/st2/Metadata on.store/default/st2/Metadata/schema on.store/default/st2/Metadata/segments on.store/default/st2/Metadata/segments/0.1_1592477533846.segment on.store/default/st2/Metadata/segments/0_1592469643544.segment on.store/default/st2/Metadata/segments/1_1592470190145.segment on.store/default/st2/Metadata/segments/2 1592470270215.segment on.store/default/st2/Metadata/segments/3 1592470293826.segment on.store/default/st2/Metadata/segments/4_1592470308901.segment on.store/default/st2/Metadata/tablestatus

The file that will get cleanup in auto cleanup

store/default/st2/Fact/PartA
n.store/default/st2/Fac1/Part0/Segment_0
n.store/default/st2/Fac1/Part0/Segment_0/0_1592469655636.carbonindexmerge
n.store/default/st2/Fac1/Part0/Segment_0/part-0-0 batchno0-0-0-1592469643544.carbondata
n.store/default/st2/Fac1/Part0/Segment_0.1
n.store/default/st2/Fact/Part0/Segment_0.1/0.1_1592477541574.carbonindexmerge
n.store/default/st2/Fact/Part0/Segment_0.1/part-0-0_batchno0-0-0.1-1592477533846.carbondata
n.store/default/st2/Fact/Part0/Segment_0.1/part-0-1_batchno0-0-0.1-1592477533846.carbondata
n.store/default/st2/Fact/Part0/Segment_1
n.store/default/st2/Fact/Part0/Segment_1/1_1592470206156.carbonindexmerge
n.store/default/st2/Fact/Part0/Segment_1/part-0-0_batchno0-0-1-1592470190145.carbondata
n.store/default/st2/Fact/Part0/Segment_2
n.store/default/st2/Fact/Part0/Segment_2/2_1592470278245.carbonindexmerge
n.store/default/st2/Fact/Part0/Segment_2/part-0-0_batchno0-0-2-1592470270215.carbondata
n.store/default/st2/Fact,Part0/Segment_2/part-0-1_batchno0-0-2-1592470270215.carbondata
n.store/default/st2/Fact/Part0/Segment_3
n.store/default/st2/Fact/Part0/Segment_3/3_1592470301948.carbonindexmerge
n.store/default/st2/Fact,Part0/Segment_3/part-0-0_batchno0-0-3-1592470293826.carbondata
n.store/default/st2/Fact/Partu/Segment_4
n.store/default/st2/Fact/Part0/Segment_4/4_1592470314996.carbonindexmerge
n.store/default/st2/Fact/Part0/Segment_4/part-0-0_batchno0-0-4-1592470308901.carbondata
n.store/default/st2/LockFiles

After auto cleanup

Cleanup the Fact/Part_x/segment folders , files .segment present in Metadata/segmets and also update tablestatus file

```
/st2/Fact/Part0/Segment 0.1
/st2/Fact/Part0/Segment_0.1/0.1_1592477541574.carbonindexmerge
/st2/Fact/Part0/Segment_0.1/part-0-0_batchno0-0-0.1-1592477533846.carbondata
/st2/Fact/Part0/Segment_0.1/part-0-1_batchno0-0-0.1-1592477533846.carbondata
/st2/Fact/Part0/Segment 4
/st2/Fact/Part0/Segment_4/4_1592470314996.carbonindexmerge
/st2/Fact/Part0/Segment 4/part-0-0 batchno0-0-4-1592470308901.carbondata
/st2/LockFiles
/st2/LockFiles/Segment 0.lock
/st2/LockFiles/Segment_1.lock
/st2/LockFiles/Segment_2.lock
/st2/LockFiles/Segment 3.lock
/st2/LockFiles/Segment_4.lock
/st2/LockFiles/clean_files.lock
/st2/LockFiles/compaction.lock
/st2/LockFiles/meta.lock
/st2/LockFiles/tablestatus.lock
/st2/LockFiles/update.lock
/st2/Metadata
/st2/Metadata/schema
/st2/Metadata/segments
/st2/Metadata/segments/0.1 1592477533846.segment
/st2/Metadata/segments/4 1592470308901.segment
/st2/Metadata/tablestatus
 st2/Metadata/tablestatus.history
```

High level design:

deleteLoadsAndUpdateMetadata using tablestatus files , finds the which are COMPACTED, MARKED_FOR_DELETE, INSERT_IN_PROGRESS, INSERT_OVERWRITE_IN_PROGRESS files and which crossed more than 60 minutes and delete them.

Auto cleanup getting triggered in case of insert

BeginTableWrite RED		MetadataManager	ClassLoaderSafeConnectorMetadata	CarbondataMetadata	CarbondataMetadataUtils	SegmentStatusManager		
IREGISTE		UNREGISTE	RED UNREGISTERED	UNREGISTERI	ED UNREGISTEI	RED UNREGIST		
1 : createWrite	rTarget	UNREGISTE	ERED UNREGISTERED	UNREGISTERI	ed unregistei	RED UNREGIST		
IRE FISTE	ERED	UNREGISTE	ERED UNREGISTERED	UNREGISTERI	ED UNREGISTEI	RED UNREGIST		
IRE GISTÊ	: beginlnse	"UNRE GISTE	ERED UNREGISTERED	UNREGISTERI	ED UNREGISTEI	RED UNREGIST		
IREGISTE		UNREGIST	BeginInsertUNREGISTERED	UNREGISTERI	ED UNREGISTEI	RED UNREGIST		
IREGISTE		UNREGISTE		UNREGISTERI	ED UNREGISTEI	RED UNREGIST		
IREGISTE		UNREGISTE	ERED UNREGISTERED	UNRE SISTERI	ED UNREGISTEI	RED UNREGIST		
IREGISTE		UNREGISTE	ERED UNREGISTERED	5 : beginInsertUpdateIn	ternal UNREGISTEI	RED UNREGIST		
IREGISTE		UNREGISTE	ERED UNREGISTERED	UNRE SISTERI	ED UNREGISTEI	RED UNREGIST		
IREGISTE		UNREGISTE	ERED UNREGISTERED	UNRE Store		RED UNREGIST		
IREGISTE		UNREGISTE	ERED UNREGISTERED	UNREGISTERI	ED UN 7; deleteLoadsAn	dUpdateMetadata EGIST		
IREGISTE		UNREGISTE	ERED UNREGISTERED	UNREGISTERI	ED UNREGISTEI	RED UN EGIST		
IREGISTE		UNREGISTE	ERED UNREGISTERED	UNREGISTERI	ED UNREGISTEI	RED UNREGIST		
IREGISTE		UNREGISTE	ERED UNREGISTERED	UNREGISTERI	ED UNREGISTEI	RED UNREGIST		
IREGISTE		UNREGISTE	ERED UNREGISTERED	UNREGISTERI	ED UNREGISTEI	RED UNREGIST		
IREGISTE	ERED	UNREGISTE	ERED UNREGISTERED	UNREGISTERI	ED UNREGISTEI	RED UNREGIST		

Auto cleanup getting triggered in case of delete

eginTableWrite	MetadataManager	ClassLoaderSafe	ConnectorMetadata	Carbonda	ataMetadata	CarbondataN	letadataUtils	SegmentStat	usManager	STEF STEF
IREGISTE	RED UNRE		UNREGIST		NREGIS	TERED		STERED		STEF
1 : createWriter	RED UNRE		UNREGIST		INREGIS	TERED		STERED		STEF
IRE DISTE	RED UNRE		UNREGIST		INREGIS	TERED		STERED		STEF
2: beginDele	etesAsInsert		UNREGIST		INREGIS	TERED		STERED		STEF
IREGISTE	RED UNSE	ginDeletesAsInsert	UNREGIST		INREGIS	TERED		STERED		STEF
IREGISTE	RED UNRE	GISTERED	4 : beginDelete	Aclacart	INREGIS	TERED		STERED		STER
IREGISTE	RED UNRE		UNREGIST	ERED	NREGIS	TERED		STERED		STER
IREGISTE	RED UNRE		UNREGIST	ER 5:beg	uinInsertUpdateInt	ernal RED		STERED		STE
IREGISTE	RED UNRE		UNREGIST		NREGIS	TERED		STERED		STE
IREGISTE	RED UNRE		UNREGIST		6 : cleanupCom	pactionTable	UNREGIS	STERED		STE
IREGISTE	RED UNRE		UNREGIST		INREGIS	TEREC] deleteLoadsAndU	IpdateMetadata		STE
IREGISTE	RED UNRE		UNREGIST		INREGIS	TERED	UNREGIS	STERE		STE
IREGISTE	RED UNRE		UNREGIST		INREGIS	TERED		STERED		STE
IREGISTE	RED UNRE		UNREGIST		INREGIS	TERED		STERED		STE
IREGISTE	RED UNRE		UNREGIST		NREGIS	TERED		STERED		STE
			UNREGIST		JNREGIS	TERED		STERED		STE
										STE

Auto cleanup getting triggered in case of **update**

BeginTableWrite	MetadataManager	ClassLoaderSafe	ConnectorMetadata	Carbonda	ataMetadata	Carbondata	NetadataUtils	SegmentStat	usManager	ERI
IREGISTE			INREGIST	FRED I	NREGIS	TERED		TERED	LINREGIST	ERI ERI
VREGISTE			UNREGIST		INREGIS		UNREGIS	TERED		ĒR
1 : createWriter	Target		UNREGIST		INREGIS		UNREGIS	TERED		ER
2 : begin			UNREGIST		INREGIS		UNREGIS	TERED		ĒR
IREGISTE		beginUpdate	UNREGIST		UNREGIS		UNREGIS	TERED		ĒR
REGISTE		SISTERED	UNREGIST	ERED U	INREGIS		UNREGIS	TERED		EF
REGISTE			4 : beginUp		INREGIS		UNREGIS	TERED		EF
REGISTE			UNREGIST	ERE5:be	L] ginInsertUpdateInt	ernal RED	UNREGIS	TERED		EF
REGISTE			UNREGIST		NREGIS		UNREGIS	TERED		EF
IREGISTE			UNREGIST		6 : cleanupCom	pactionTable	UNREGIS	TERED		ĒF
IREGISTE			UNREGIST		UNREGIS	TERED	_ deleteLoadsAndU	odateMetadata		ĒF
IREGISTE			UNREGIST		UNREGIS		UNREGIS	TERE	UNREGIST	EF
IREGISTE	RED UNREG		UNREGIST		UNREGIS		UNREGIS	TERED	UNREGIST	EF
IREGISTE	RED UNREG		UNREGIST		UNREGIS		UNREGIS	TERED		EF
REGISTE	RED UNREG		UNREGIST		UNREGIS		UNREGIS	TERED		ĒF
										EF
IREGISTE	RED UNREG	SISTERED	UNREGIST	ERED I	UNREGIS	TERED	UNREGIS	TERED	UNREGIST	ĒF

Auto cleanup getting triggered in case of vacuum

BeginTableWrite	MetadataManager	ClassLoaderSafe	ConnectorMetadata	Carbond	lataMetadata	Carbonda	taMetadataUtils	SegmentSta	tusManager
NREGISTE	RED UNREG		UNREGIST		UNREGI		D UNREG	ISTERED	
1 : createWriter	RED UNREO		UNREGIST		UNREG		D UNREG	ISTERED	
NREFISTE	RED UNREG		UNREGIST		UNREGI		D UNREG	ISTERED	
			UNREGIST		UNREGI		D UNREG	ISTERED	
REGISTE	RED UNREG	beginVacuum ED	UNREGIST		UNREGI		D UNREG	ISTERED	
REGISTE	RED UNREG	GISTERED	4 : beginVa	ERED	UNREGI		D UNREG	ISTERED	
REGISTE	RED UNREG	GISTERED	UNREGIST	ERED	INREGI		D UNREG	ISTERED	
REGISTE	RED UNREG		UNREGIST	ERE5:be	ginInsertUpdate	eInternal R =	D UNREG	ISTERED	
REGISTE	RED UNREG		UNREGIST		INREG	STERE	D UNREG	ISTERED	
REGISTE	RED UNREG		UNREGIST		6 : cleanupC	ompactionTabl	UNREG	ISTERED	
REGISTE	RED UNREG		UNREGIST		UNREGI		7 : deleteLoadsAn	dUpdateMetadata	
REGISTE	RED UNREG		UNREGIST		UNREGI		D UNREG	ISTERE	UNREG
IREGISTE	RED UNREG		UNREGIST		UNREGI		D UNREG	ISTERED	
IREGISTE	RED UNREG		UNREGIST		UNREGI		D UNREG	ISTERED	
IREGISTE	RED UNREG		UNREGIST		UNREGI		D UNREG	ISTERED	
REGISTE	RED UNREG	GISTERED	UNREGIST	ERED	UNREGI	STERE	D UNREG	ISTERED	UNREG