Auto Clean Up for Hive ORC

Since, openlookeng is now supporting compaction/vacuum operation on Hive transactional tables which merges all base and delta files together (according to the compaction type). So that upcoming read queries can be redirected to use newly created, compacted directories which improves performance. This makes old, smaller delta directories eligible for cleanup. But cleaner operation has to wait for any read operation started before vacuum operation since it might still be using old delta directories.

Therefore, a background process is required which will be responsible for removing any redundant directory from table path.

Directory structure of table `table_txn_001` after 3 insert operations:

drwxrwxrwx	- roo	ot supergroup	Θ	2020-07-09	20:50)/user/hive/warehouse/table txn 001/delta 0000001 0000001 00000
- rw- r r	3 roc	ot supergroup	1	2020-07-09	20:50	<pre>/user/hive/warehouse/table_txn_001/delta_0000001_000001_0000/_orc_acid_version</pre>
- rw-rr	3 roc	ot supergroup	901	2020-07-09	20:50) /user/hive/warehouse/table_txn_001/delta_0000001_0000001_0000/bucket_00000
drwxrwxrwx	- roo	ot supergroup	Θ	2020-07-09	20:50) /user/hive/warehouse/table_txn_001/delta_0000002_0000002_0000
- rw- r r	3 roc	ot supergroup	1	2020-07-09	20:50	<pre>//user/hive/warehouse/table_txn_001/delta_0000002_000002_0000/_orc_acid_version</pre>
- rw- r r	3 roc	ot supergroup	909	2020-07-09	20:50) /user/hive/warehouse/table_txn_001/delta_0000002_0000002_0000/bucket_00000
drwxrwxrwx	- roc	ot supergroup	Θ	2020-07-09	20:50)/user/hive/warehouse/table_txn_001/delta_0000003_0000003_0000
- rw-rr	3 roc	ot supergroup	1	2020-07-09	20:50	<pre>/user/hive/warehouse/table_txn_001/delta_0000003_0000003_00000/_orc_acid_version</pre>
- rw- r r	3 roc	ot supergroup	905	2020-07-09	20:50) /user/hive/warehouse/table txn 001/delta 0000003 0000003 0000/bucket 00000

After minor compaction on table `table_txn_001`

Command:

VACUUM TABLE table_txn_001;

	root	supergroup	0	2020-07-09	20:50	/user/hive/warehouse/table txn 001/delta 0000001 0000001 0000
3	root	supergroup	1	2020-07-09	20:50	/user/hive/warehouse/table_txn_001/delta_0000001_000001_0000/_orc_acid_version
3	root	supergroup	901	2020-07-09	20:50	/user/hive/warehouse/table_txn_001/delta_0000001_0000001_0000/bucket_00000
	root	supergroup	Θ	2020-07-09	20:55	/user/hive/warehouse/table_txn_001/delta_0000001_0000003
3	root	supergroup	1	2020-07-09	20:55	/user/hive/warehouse/table_txn_001/delta_0000001_0000003/_orc_acid_version
3	root	supergroup	916	2020-07-09	20:55	/user/hive/warehouse/table_txn_001/delta_0000001_0000003/bucket_00000
	root	supergroup	Θ	2020-07-09	20:50	/user/hive/warehouse/table_txn_001/delta_0000002_0000002_0000
3	root	supergroup	1	2020-07-09	20:50	/user/hive/warehouse/table_txn_001/delta_0000002_0000002_0000/_orc_acid_version
3	root	supergroup	909	2020-07-09	20:50	/user/hive/warehouse/table_txn_001/delta_0000002_0000002_0000/bucket_00000
	root	supergroup	Θ	2020-07-09	20:50	/user/hive/warehouse/table_txn_001/delta_0000003_0000003_0000
3	root	supergroup	1	2020-07-09	20:50	/user/hive/warehouse/table_txn_001/delta_0000003_000003_0000/_orc_acid_version
3	root	supergroup	905	2020-07-09	20:50	/user/hive/warehouse/table_txn_001/delta_0000003_0000003_0000/bucket_00000
	• • • • • • • • • • • • • • • • •	 root 	 root supergroup 	 root supergroup 3 root supergroup 3 root supergroup 3 root supergroup 905 	- root supergroup 0 2020-07-09 3 root supergroup 1 2020-07-09 3 root supergroup 901 2020-07-09 - root supergroup 0 2020-07-09 3 root supergroup 1 2020-07-09 3 root supergroup 1 2020-07-09 3 root supergroup 1 2020-07-09 3 root supergroup 916 2020-07-09 - root supergroup 0 2020-07-09 3 root supergroup 1 2020-07-09 3 root supergroup 1 2020-07-09 3 root supergroup 905 2020-07-09	- root supergroup 0 2020-07-09 20:50 3 root supergroup 1 2020-07-09 20:50 3 root supergroup 901 2020-07-09 20:50 - root supergroup 0 2020-07-09 20:55 3 root supergroup 1 2020-07-09 20:55 3 root supergroup 1 2020-07-09 20:55 3 root supergroup 0 2020-07-09 20:55 - root supergroup 0 2020-07-09 20:55 3 root supergroup 0 2020-07-09 20:50 3 root supergroup 0 2020-07-09 20:50 - root supergroup 0 2020-07-09 20:50 3 root supergroup 0 2020-07-09 20:50 - root supergroup 0 2020-07-09 20:50 3 root supergroup 0 2020-07-09 20:50

Directory created by Vacuum:

delta_0000001_0000003

Directories ready for clean up:

delta_000001_000001_0000

delta_000002_000002_0000

delta_000003_000003_0000

After major compaction on table `table_txn_001`

Command:

VACUUM TABLE table_txn_001 FULL;

drwxrwxrwx	- root	supergroup	Θ	2020-07-09	20:57	/user/hive/warehouse/table_txn_001/base_0000003
-rw-rr	3 root	supergroup	1	2020-07-09	20:57	/user/hive/warehouse/table_txn_001/base_0000003/_orc_acid_version
-rw-rr	3 root	supergroup	916	2020-07-09	20:57	/user/hive/warehouse/table_txn_001/base_0000003/bucket_00000
drwxrwxrwx	- root	supergroup	Θ	2020-07-09	20:56	/user/hive/warehouse/table_txn_001/delta_0000001_0000001_0000
- rw-rr	3 root	supergroup	1	2020-07-09	20:56	/user/hive/warehouse/table_txn_001/delta_0000001_000001_0000/_orc_acid_version
- rw-rr	3 root	supergroup	903	2020-07-09	20:56	/user/hive/warehouse/table_txn_001/delta_0000001_0000001_0000/bucket_00000
drwxrwxrwx	- root	supergroup	Θ	2020-07-09	20:56	/user/hive/warehouse/table_txn_001/delta_0000002_0000002_0000
-rw-rr	3 root	supergroup	1	2020-07-09	20:56	/user/hive/warehouse/table_txn_001/delta_0000002_0000002_0000/_orc_acid_version
- rw-rr	3 root	supergroup	910	2020-07-09	20:56	<pre>/user/hive/warehouse/table_txn_001/delta_0000002_0000002_0000/bucket_00000</pre>
drwxrwxrwx	- root	supergroup	Θ	2020-07-09	20:56	/user/hive/warehouse/table_txn_001/delta_0000003_000003_0000
- rw-rr	3 root	supergroup	1	2020-07-09	20:56	/user/hive/warehouse/table_txn_001/delta_0000003_000003_0000/_orc_acid_version
-rw-rr	3 root	supergroup	905	2020-07-09	20:56	/user/hive/warehouse/table txn 001/delta 0000003 0000003 0000/bucket 00000

Directory created by Vacuum:

base_0000003

Directories ready for clean up:

delta_0000001_000001_0000

delta_000002_000002_0000

delta_000003_000003_0000

After major compaction followed by minor compaction on table `table_txn_001`

drwxrwxrwx		root	supergroup	Θ	2020-07-09	21:21	/user/hive/warehouse/table_txn_001/base_0000003
- rw-rr	3	root	supergroup	1	2020-07-09	21:21	/user/hive/warehouse/table_txn_001/base_0000003/_orc_acid_version
- rw-rr	3	root	supergroup	917	2020-07-09	21:21	/user/hive/warehouse/table_txn_001/base_0000003/bucket_00000
drwxrwxrwx		root	supergroup	Θ	2020-07-09	21:19	/user/hive/warehouse/table_txn_001/delta_0000001_0000001_0000
- rw-rr	3	root	supergroup	1	2020-07-09	21:19	/user/hive/warehouse/table_txn_001/delta_0000001_000001_0000/_orc_acid_version
- rw-rr	3	root	supergroup	913	2020-07-09	21:19	/user/hive/warehouse/table_txn_001/delta_0000001_0000001_0000/bucket_00000
drwxrwxrwx		root	supergroup	Θ	2020-07-09	21:21	/user/hive/warehouse/table_txn_001/delta_0000001_0000003
-rw-rr	3	root	supergroup	1	2020-07-09	21:21	/user/hive/warehouse/table_txn_001/delta_0000001_0000003/_orc_acid_version
-rw-rr	3	root	supergroup	918	2020-07-09	21:21	/user/hive/warehouse/table_txn_001/delta_0000001_0000003/bucket_00000
drwxrwxrwx		root	supergroup	Θ	2020-07-09	21:19	/user/hive/warehouse/table_txn_001/delta_0000002_0000002_0000
-rw-rr	3	root	supergroup	1	2020-07-09	21:19	/user/hive/warehouse/table_txn_001/delta_0000002_0000002_0000/_orc_acid_version
- rw-rr	3	root	supergroup	903	2020-07-09	21:19	/user/hive/warehouse/table_txn_001/delta_0000002_0000002_0000/bucket_00000
drwxrwxrwx		root	supergroup	Θ	2020-07-09	21:20	/user/hive/warehouse/table_txn_001/delta_0000003_000003_0000
-rw-rr	3	root	supergroup	1	2020-07-09	21:20	/user/hive/warehouse/table_txn_001/delta_0000003_000003_0000/_orc_acid_version
- rw-rr	3	root	supergroup	910	2020-07-09	21:20	/user/hive/warehouse/table txn 001/delta 0000003 0000003 0000/bucket 00000

Directory created by last Vacuum:

base_0000003

Directories ready for clean up:

delta_0000001_0000003

delta_0000001_000001_0000

delta_000002_000002_0000

delta_000003_000003_0000

> After minor compaction on table with delete delta directories

drwxrwxrwx		root	supergroup	Θ	2020-07-09	21:23	23 /user/hive/warehouse/table_txn_001/delete_delta_0000001_0000004
- rw-rr	3	root	supergroup	1	2020-07-09	21:23	23 /user/hive/warehouse/table_txn_001/delete_delta_0000001_0000004/_orc_acid_version
- rw-rr	3	root	supergroup	786	2020-07-09	21:23	23 /user/hive/warehouse/table txn 001/delete delta 0000001 0000004/bucket 00000
drwxrwxrwx		root	supergroup	Θ	2020-07-09	21:22	22 /user/hive/warehouse/table_txn_001/delete_delta_0000004_0000004_00000
- rw-rr	3	root	supergroup	1	2020-07-09	21:22	22 /user/hive/warehouse/table_txn_001/delete_delta_0000004_000004_00000/ orc acid version
-rw-rr	3	root	supergroup	786	2020-07-09	21:22	22 /user/hive/warehouse/table_txn_001/delete_delta_0000004_000004_0000/bucket 00000
drwxrwxrwx		root	supergroup	Θ	2020-07-09	21:22	22 /user/hive/warehouse/table_txn_001/delta_0000001_0000001_0000
-rw-rr	3	root	supergroup	1	2020-07-09	21:22	22 /user/hive/warehouse/table_txn_001/delta_0000001_0000001_0000/ orc acid version
- rw-rr	3	root	supergroup	907	2020-07-09	21:22	22 /user/hive/warehouse/table_txn_001/delta_0000001_0000001_0000/bucket 00000
drwxrwxrwx		root	supergroup	Θ	2020-07-09	21:23	23 /user/hive/warehouse/table_txn_001/delta_0000001_0000004
- rw-rr	3	root	supergroup	1	2020-07-09	21:23	23 /user/hive/warehouse/table_txn_001/delta_0000001_0000004/_orc_acid_version
- rw-rr	3	root	supergroup	938	2020-07-09	21:23	23 /user/hive/warehouse/table_txn_001/delta_0000001_0000004/bucket 00000
drwxrwxrwx		root	supergroup	Θ	2020-07-09	21:22	22 /user/hive/warehouse/table_txn_001/delta_0000002_0000002_00000
- rw-rr	3	root	supergroup	1	2020-07-09	21:22	22 /user/hive/warehouse/table txn 001/delta 0000002 0000002 0000/ orc acid version
- rw-rr	3	root	supergroup	911	2020-07-09	21:22	22 /user/hive/warehouse/table_txn_001/delta_0000002_0000002_0000/bucket 00000
drwxrwxrwx		root	supergroup	Θ	2020-07-09	21:22	22 /user/hive/warehouse/table_txn_001/delta_0000003_0000003_00000
- rw-rr	3	root	supergroup	1	2020-07-09	21:22	22 /user/hive/warehouse/table_txn_001/delta_0000003_0000003_0000/ orc acid version
- rw-rr	3	root	supergroup	908	2020-07-09	21:22	22 /user/hive/warehouse/table_txn_001/delta_0000003_0000003_0000/bucket_00000
drwxrwxrwx		root	supergroup	Θ	2020-07-09	21:22	22 /user/hive/warehouse/table_txn_001/delta_0000004_000004_00000
- rw-rr	3	root	supergroup	1	2020-07-09	21:22	22 /user/hive/warehouse/table_txn_001/delta_0000004_000004_00000/_orc_acid_version
- rw-rr	3	root	supergroup	925	2020-07-09	21:22	22 /user/hive/warehouse/table_txn_001/delta_0000004_000004_00000/bucket_00000

Directory created by last Vacuum:

delta_0000001_0000004

delete_delta_0000001_0000004

Directories ready for clean up:

delta_0000001_000001_0000

delta_000002_000002_0000

delta_000003_000003_0000

delta_0000004_000004_0000

delete_delta_0000004_000004_0000

Approach:

To perform cleanup after vacuum, CleanerTask will be submitted to ScheduledExecutorService from HiveMetadata#commit(). Till this stage, new directory is created and old/compacted directories can be cleaned. But before cleanup we have to wait for finish of any readers using old/compacted directories.

CleanerTask#readyToClean() is used to insure that each read lock taken on that table before compaction must finish. This function makes call to ThriftHiveMetastoreClient#showLocks to get a list of read locks on table. This function returns true if all locks taken before compaction, are released. Otherwise returns false.

On receiving false from readyToClean(), CleanerTask will stop current thread execution and wait for ExecutorService to schedule CleanerTask again, that would be after configurable value (hive.compaction-cleanup-interval).

True from readyToClean() signals CleanerTask to start remove files. List of directories which have to be removed will be received from AcidUtils.Directory.getObsolete().

- After each vacuum commit, a CleanerTask will be submitted to scheduled executor service for cleaning obsolete files.
- Configuration added:
 - *hive.compaction-cleanup-interval* : (Milis) Interval after which cleanup task will be resubmitted if there is any unreleased read lock on table. Default value is 2000.
 - *hive-compaction-cleanup-threads* : Number of threads to run in compaction cleanup service.
 Default value is 2.

1 : finishVa 4 : com	cuum 8 exec	create» utorService, Clea				UNREG			STERED I	INREGIS		INREG
4 : com	nit	utorService, Clea	ner D U						ALCUED A			ALLINE
4 : com			ED U			UNREG			STERED I	JNREGIS		UNREG
GISTER	אַרן לישא	DECIST				UNREG			STERED I	JNREGIS		UNREG
dieter		1101011	RED U	NREGIS	5 : commit	UNREG	STERE) UNREGI	STERED I	JNREGIS		UNRE
GISTER	REDUN	6 : start	create»	CleanerSC	leanerTack	INREG	STERE) UNREGI	TERED I	JNREGIS		UNRE
EGISTEF	RED UN	REGIST	RED U		Jeanerrask	interval	aneriask, –		STERED I	JNREGIS		UNRE
	RED UN	REGIST	8 : schedul	eAtFixedRate		INREG	Execute		STERED I	JNREGIS		UNRE
	RED UN	REGISTE	RED U			JNREG	rate untill		STERED I	JNREGIS		UNRE
		REGISTE	RED U	[9 :	WNREG	explicitly	LINREGI	STERED I	JNREGIS		UNRE
		REGISTE		ns true if a	10 : rea	adyToClean	STERED	table and	STERED I	JNREGIS		UNRE
		REGISTE	REC locks	sare sed		11 : showLocks	STERE	UNREG	STERED I	JNREGIS		UNRE
		REGISTE	REDU			JNREG			12; sh	iowLocks	TERED.	NRE
		REGISTE	RED U		13:retu	im if false	to start clean	up task again	STERED I	JNREGIS		UNRE
		REGISTE	RED U		14 : rer	noveFiles EC	15 : getAcidSt	UNREGI	STERED I	JNREGIS		
		REGISTE	RED U			JNREG	STEREL	UNREGI	STERED (JNRI GIS		
			move directori	es listed by		UNREG	S 16 : dire	ctory	STERED 1	JNREGIS		
		REGIS	ectory,getObso	lete		JNREG	STERE) ÚNREGI	STERED I	JNREGIS		
		REGIST	RED ¹ 7	stop 1	g: cancelSched	luledService						
		REGISTE	RED U			JNREG	STERED					

Flow diagram: