Data purging concepts

1. Referenced expired rows from child tables must be deleted before rows from parent table can be deleted. Child tables are defined by a Foreign Key constraint.

For example, in database OSM child tables Entity, Event\_Activity, Event\_Fill have FK to parent table Entity\_Summary.

1. Data expiration must be defined: if based on a row’s timestamp in days – 7 days or 2 weeks or 1 month etc.
2. Rows from child tables are deleted based on expired IDs (purge keys) from a parent table and then from parent table.
3. Data must be deleted in batches of rows until the purge target is met.

For example, 100000 rows in a batch size.

1. Encapsulating data purge from all involved tables in 1 transaction presents a potential blocking/deadlocking issue for other processes including a mandatory maintenance system process to backup transaction log.
2. Errors during data purge must be captured and optionally persisted.
3. Data purge statistics must be recorded in a log.
4. Data purge process must have a run time limit.
5. Data purge must be re-entrant to continue from the point of interruption either if its process was killed, stopped due to exception or exceeded a run time limit.

In summary, to be a robust solution all these points must be implemented.