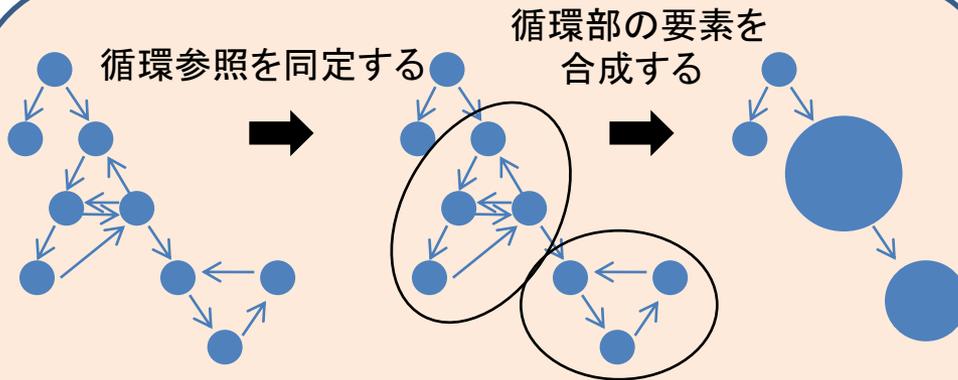


オープンかつポータブルなデータベース garbage collection データベースの不要なデータを同定・削除する

郷原浩之 東京大学大学院工学系研究科システム創成学専攻

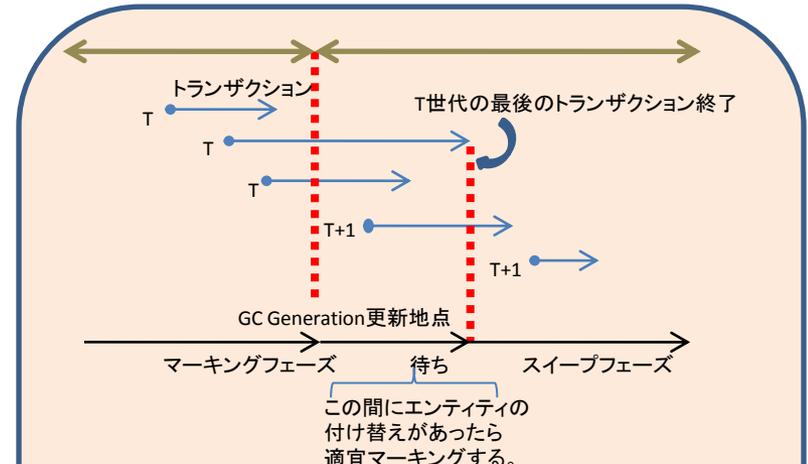
データベース内には決して使われないことが保証できるようなデータが存在している。しかしながら安易なデータの削除はDBの破壊を招くため、通常データは削除せず残しておく。一方で、情報流出などのリスク故に不必要なデータを削除したいニーズは大きい。本ソフトウェアでは、このような使用されることがないデータを自動的に同定し、削除する機能を提供する。

マーキングはSQLで行う。
SQLの発行回数を最小にするための工夫



本ソフトウェアではマークアンドスイープ法を用いて garbage collection を行う。SQL を用いてマーキングを行う。SQL の発行はディスクへのアクセスが伴うため、その回数を最小化したい。そのためにエンティティの参照関係を解析し、最適な SQL の発行順序を決定する。

スイープを任意のタイミングで行える
湯浅式 GC を DB で実現するための工夫



湯浅式 GC に GC 世代の概念を持ち込みトランザクションが実行されていてもデータベースの garbage collection が可能となった。DBGCC に世代の概念を持ち込むことで、スイープは任意のタイミングで行える。