大学共同利用機関法人 情報・システム研究機構
国立情報学研究所
National Institute of Informatics

アドバンス・トップエスイー プロフェッショナルスタディ

形式手法を用いたストレージレイヤOSSの信頼性評価の試み

株式会社NTTデータ　　関　堅吾

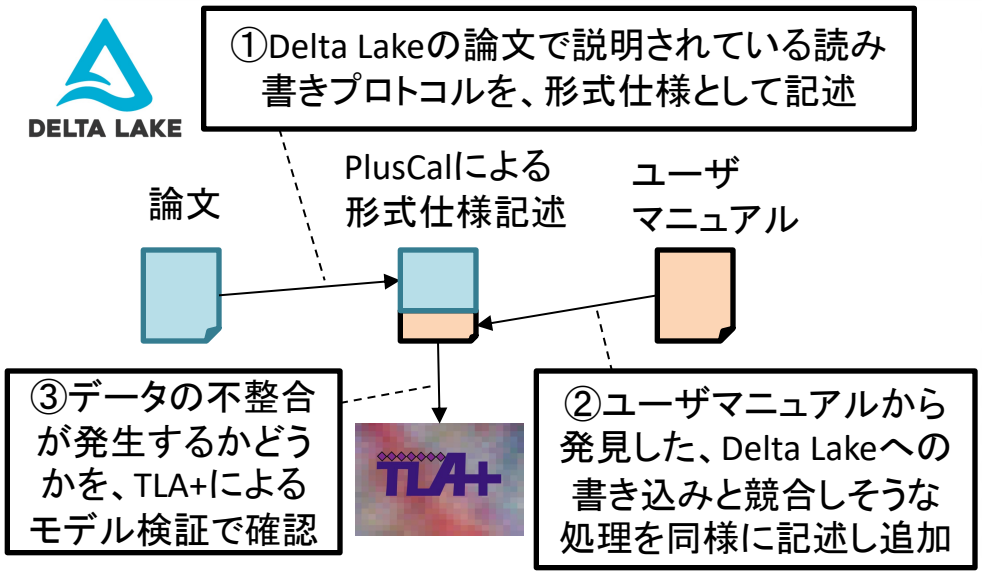
OSSの品質検証における問題点

第三者が開発したビッグデータ関連の**OSSを業務で扱うことが多いが、異常系のテストの実施方針や結果に確信が持てない**。現状では、システムの構成要素を順番に故障させていき、その影響や復旧方法を確認することが多いが、**たまたまそのときの状態やタイミングによって観測した結果になったのではないかという疑念を払拭できない**。

手法・ツールの適用による解決

複数のビッグデータ関連OSSで採用実績のある形式言語・ツールであるTLA+を用いて、ストレージレイヤOSS(※)であるDelta LakeとApache Icebergを題材に**データの不整合が発生するケースを発見**。一般的に形式手法はこれから開発するソフトウェアの仕様の妥当性を確認するために使われることが多いが、**第三者が開発したOSSの品質検証にも形式手法が活用できる**ことを示した。

取り組みの概要



結果

TLA+によるモデル検証によって、以下の問題を発見した。

Delta Lake: 書き込み中の特定タイミング（実データ書き込み後、メタデータ書き込み前）に、競合する処理(VACUUMと呼ばれる、メタデータから参照されていない実データを削除する処理) が実行されると、データの不整合が発生する可能性がある。

Iceberg: メタデータの書き込み中にネットワーク障害が発生すると、クライアントとサーバ間で状態の認識が食い違った結果、必要なデータが誤って削除される場合がある。

結果的にはいずれも既知の問題であり、ワークアラウンドが存在するか既に改修済みであったが、**発生させるのが難しいコーナーケースを形式手法で発見することができた**。

今後の展望・課題

■実際のテスト設計業務における本手法の活用

- 検出が難しいコーナーケースの発見や、テスト仕様書の品質向上を実業務で達成する
- バグ検出数、作成したテストケース数、確認観点や内容の厳密さ、テスト設計に要した期間などで効果を定量的に評価する

■形式仕様記述を正しく・効率よく作成する方法の探求

- 検証したい仕様に対応するソースコードをどのように発見するか
- 発見したソースコードのうち、どの部分を残してどの程度抽象化するか
- 作成した仕様記述が実装と一致していることをどのように保証するか

※ストレージレイヤOSS：分散ファイルシステムやクラウド上のオブジェクトストレージに格納された多数のファイル群をテーブルとして捉え、クライアントにACIDトランザクションや読み込みの高速化を提供するミドルウェアOSS。