

自己結合SQLクエリ検出ツール によるチューニングの提案

株式会社インサイトテクノロジー

岸本 拓也

tkishimo@insight-tec.co.jp

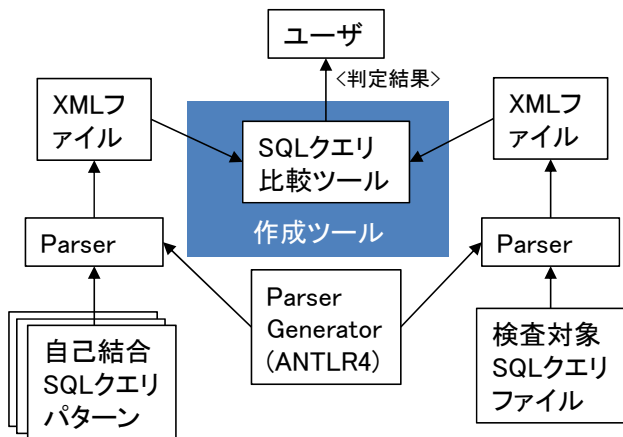
負荷テストの問題点

エンタープライズ向けの開発案件におけるパフォーマンス要件は、負荷テストによって妥当性確認が実施されていることが多い。開発最終工程で実施される負荷テストでパフォーマンス問題に対応すると手戻りコストが大きくなってしまふ。単体・結合テスト段階でパフォーマンス問題があるSQLクエリを抽出できないか検討した。

ツールの適用による解決

同一表に複数回アクセスしている自己結合SQLクエリは、1アクセスによるSQLクエリに書換えられるパターンがあることが経験上分かっていた。本研究にてパターンマッチングによる自己結合SQLクエリ検出ツールを開発した。これにより、小規模データで実施される単体・結合テストでもパフォーマンスチューニングが可能となった。

ツールのアーキテクチャ



SQLクエリパターンと書換え例

<<SQLクエリパターンファイル例 ※実行時間例(3.94秒)>>

```
SELECT col1,col2,col3 FROM tab1 a, (SELECT col1,MAX(col2) col2 FROM tab1 GROUP BY col1) b where a.col1=b.col1 and b.col2=b.col2
```

自己結合SQLクエリは書換えることで応答時間を短縮可能

<<SQLクエリパターン書換え例 ※実行時間例(2.66秒)>>

```
SELECT col1,col2,col3 FROM (select col1,col2,row_number() over(partition by col1 order by col2 desc) row_number from tab1) where row_number=1
```

Window関数の利用によるチューニング

比較ツール機能概要

特徴1: パターンの自動取り込み

パターンファイルはSQLクエリファイルとして定義。ツールが取り込む

特徴2: リスト要素による比較

構文解析結果をリスト型に変換し、比較

In

<<自己結合SQLクエリパターン>>

```
select col1,col2 from tab1 a, (select col1,max(col2) from tab1 group by col1) b where a.col1=b.col1 and a.col2=b.col2
```

<<検査対象SQLクエリ>>

```
select trim(to_char(a.val1,'00000000000000')) ,1.trim(to_char(a.val2,'00000000000000')) ,a.val3 ,a.val4 ,a.val5 ,a.val6.
```

抽象化

1. XMLファイルの'タグ'と'要素'をリスト型に変換
2. TOKENタグを含むリストを抽出
3. 表・列名などの個別情報を特定の名前に変換
4. 突合に不要なタグをリストから削除

Out

```
['', 'parse', 'sql_stmt_list', 'sql_stmt', 'factored_select_stmt', 'select_core', 'K_SELECT', 'TOKEN', 'select']
```

```
['', 'parse', 'sql_stmt_list', 'sql_stmt', 'factored_select_stmt', 'select_core', 'result_column', 'column_name', 'any_name', 'IDENTIFIER', 'TOKEN'] ..
```

比較

```
['', 'parse', 'sql_stmt_list', 'sql_stmt', 'factored_select_stmt', 'select_core', 'K_SELECT', 'TOKEN', 'select']
```

```
['', 'parse', 'sql_stmt_list', 'sql_stmt', 'factored_select_stmt', 'select_core', 'result_column', 'function_name', 'any_name', ...
```

実証実験結果

作成した9パターンが自己結合SQLクエリを抽出できるか検証

※商用環境データベースから実験対象SQLクエリを抽出し、評価

	自己結合クエリ	非自己結合クエリ
検査対象件数	131	261
検出できた件数	129	-
誤検出件数	2	30

ツールのパターンマッチは再現率98%、精度91%

クエリ構造のパターンマッチにより、負荷テストに依らずにパフォーマンス改善可能なSQLクエリの抽出ができた