

Восьмая независимая
научно-практическая конференция
«Разработка ПО 2012»

1 - 2 ноября, Москва



Автоматизация миграции динамически формируемых
запросов



Семён Григорьев
Яков Кириленко



Реинжиниринг БД

- MS SQL Server T-SQL и Oracle PL-SQL — два различных языка.
- 2,7 MLOC хранимых процедур
- 7000 процедур
- 3000 динамических EXECUTE (1500 сложных)
- От 7 до 212 операторов на литерал
- Около 40 в среднем
- 1 ч/д на сложный EXECUTE = 10 человеко-лет.
- Менее года на проект (на все работы)
- Только автоматизация.

Пример

- **Исходный код на T-SQL:**

```
IF @X = @Y
    SET @TABLE = '#table1'
ELSE
    SET @TABLE = 'table2'
SET @S = 'SELECT x FROM' + @TABLE + ' WHERE ISNULL(n,0) > 1'
EXECUTE (@S)
```

- **Результирующий код на PL-SQL:**

```
IF lv_X = lv_Y
    lv_TABLE := '#table1';
ELSE
    lv_TABLE := 'table2';
END IF;
lv_S := 'SELECT x FROM ' || lv_TABLE || ' WHERE ISNULL(n,0) > 1';
EXECUTE IMMEDIATE lv_S;
```

Варианты решения

- Динамика
 - Трансляция во время выполнения.
 - Реализация функции трансляции и встраивание её вызовов в нужные места.
- Статика
 - Вычислить новые значения для переменных, участвующих в построении запроса, на этапе трансляции.

Результат динамики

- **Исходный код на T-SQL:**

```
IF @X = @Y
    SET @TABLE = '#table1'
ELSE
    SET @TABLE = 'table2'
SET @S = 'SELECT x FROM ' + @TABLE + ' WHERE ISNULL(n,0) > 1'
EXECUTE (@S)
```

- **Результирующий код на PL-SQL:**

```
IF lv_X = lv_Y
    lv_TABLE := '#table1';
ELSE
    lv_TABLE := 'table2';
END IF;
lv_S := 'SELECT x FROM ' || lv_TABLE || ' WHERE ISNULL(n,0) > 1';
CUSTOM_EXECUTE_IMMEDIATE(lv_S);
```

Динамика. Оценка решения.

- Плюсы
 - Можно переиспользовать основной транслятор.
 - Качество результата = качество транслятора.
 - Качество транслятора можно проверить.
Иногда даже формально.
- Минусы
 - Уменьшается производительность.
 - Ухудшается сопровождаемость.
 - Уменьшается надёжность целевой системы.

Статика

- Динамические запосы статически транслируются в целевой язык.
 - Для всех переменных, участвующих в формировании запроса, новые значения вычисляются на этапе трансляции.
 - Значения вычисляются так, чтобы из переменных строился запрос на целевом языке, эквивалентный исходному.

Результат статистики

- Исходный код на T-SQL:

```
IF @X = @Y
    SET @TABLE = '#table1'
ELSE
    SET @TABLE = 'table2'
SET @S = 'SELECT x FROM ' + @TABLE + ' WHERE ISNULL(n,0) > 1'
EXECUTE (@S)
```

- Результирующий код на PL-SQL:

```
IF lv_X = lv_Y
    lv_TABLE := 'tt_table1';
ELSE
    lv_TABLE := 'new_table2';
END IF;
lv_S := 'SELECT new_x FROM ' || lv_TABLE || ' WHERE NVL(n,0) > 1';
OPEN new_cursor FOR lv_S;
```


Статика. Оценка решения.

- Плюсы
 - Весь код в едином стиле, на одном языке.
 - Нет дополнительных проблем с производительностью.
- Минусы
 - Задача в общем случае не решается.
 - Появляются эвристики.
 - Проверка корректности усложняется.
 - Ограничение на языки: синтаксис исходного и целевого языков должен быть структурно похож для семантически одинаковых конструкций.
 - Решение требует больших вычислительных ресурсов.

Решение (1)

- Смешанный подход:
 - Основа — статика.
 - Там где статика не справляется — динамика.
- Динамика применяется для конструкций, требующих сложных структурных преобразований при трансляции.
- Динамика совмещена со статикой.
 - Статическая трансляция с добавлением дополнительной информации в литералы.
 - Окончательное построение запроса во время выполнения.

Решение (2)

- Основа статистики: абстрактный синтаксический анализ.
- Уменьшение требуемых ресурсов.
 - Загрубления в представлении множества всех возможных запросов.
 - Фильтрация леса во время синтаксического анализа.
- Обеспечение качества результата.
 - Валидация результатов трансляции динамических запросов.
 - Автоматизированное тестирование результатов трансляции.

Результаты

- Полностью автоматически корректно переведено 50% запросов.
- После доработки эвристик под конкретную задачу автоматически переведено 72% запросов.
- Остальные требуют дальнейшей ручной доработки.
- 98% запросов обрабатывается за разумное время (ночное тестирование).
- Статический анализ запросов позволил выделить запасы, возвращающие результат.
 - Скорректировать способ выполнения запроса.
 - Скорректировать сигнатуры процедур.

Оценка опыта

- Формально корректным может быть только смешанное решение (АОТ + JIT)
- Статика с валидацией и динамическим тестированием — приемлемое инженерное решение.

Возможное использование

- Пригодится для
 - Встраиваемые языки.
 - Макропроцессоры (M4, make, GNU C)
 - Eval (JavaScript, динамический SQL)
- Пригодится при
 - Миграции (трансляции)
 - Рефакторингах
 - JIT/AOT