

# Статический анализатор SparkSQL

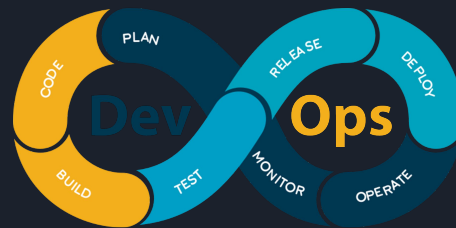
Команда "Кукумбер"

Харламов Александр  
Соломатин Павел

DataHack 2022

# Цель

Упростить рефакторинг кода и CI/CD процессы при деплое приложений с использованием SQL скриптов.





## Решение

CLI-программа реализует следующие виды анализа:

- синтаксический
- стилистический
- семантический

Есть возможность легко добавлять правила.

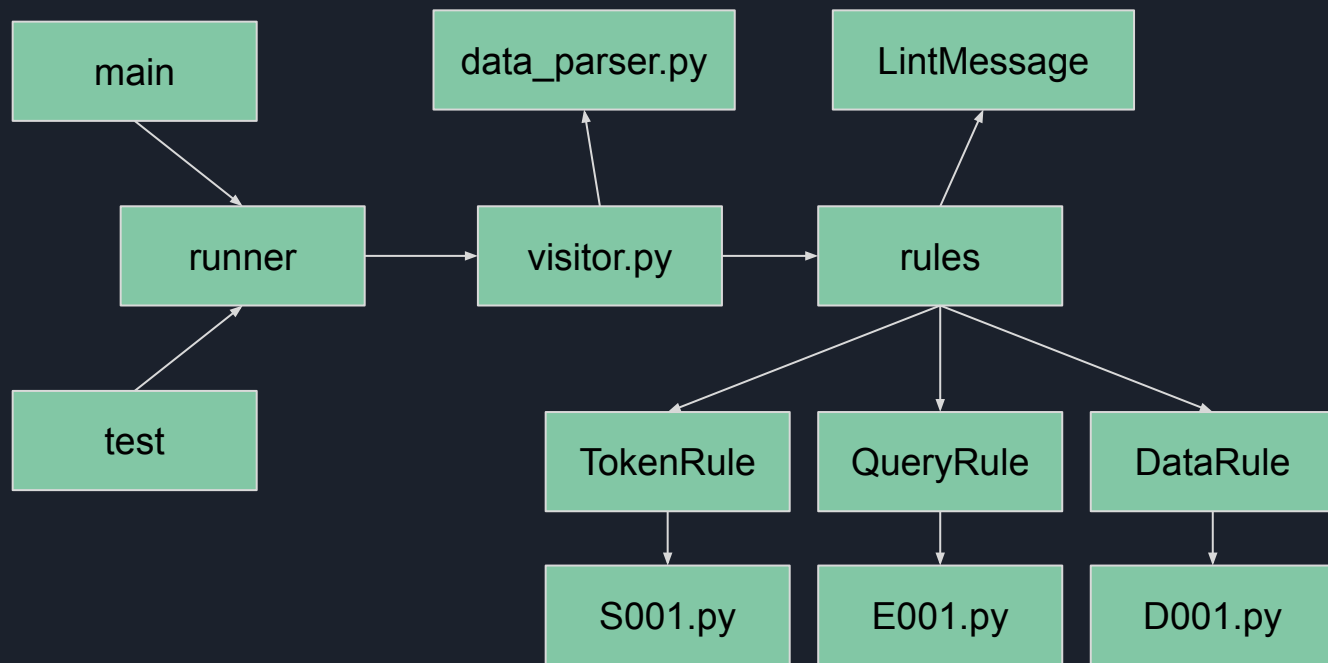


# Стек технологий

- python            Основной язык
- sqlparse        Парсинг SQL-кода
- inflection       Трансформация строк



# Архитектура





# Фичи

- Правила - классы
- При наследовании нового правила оно добавляется в общий список
- Правила могут касаться как одного слова, так и запроса целиком
- При обработке запроса CREATE TABLE запоминается структура таблицы
- Интерфейс командной строки



# Фичи

```
class TokenRule(Rule, abc.ABC):  
    """Правило для одного токена"""  
    def __init_subclass__(cls, **kwargs):  
        super().__init_subclass__(**kwargs)  
        token_rules.append(cls())  
  
    @abc.abstractmethod  
    def is_suitable(self, obj: sql.Token) -> bool:  
        """Подходит ли по шаблону"""  
        raise NotImplemented  
  
    @abc.abstractmethod  
    def is_correct(self, obj: sql.Token) -> bool:  
        """Выполняется ли правило"""  
        raise NotImplemented
```

Интерфейс правила



# Пример работы (входной файл)

```
SELECT ;

SELECT FROM;

SELECT a FROM;

SELECT from b;

SELECT Id FROM (SELECT id, clientNames
                  FROM clients WHERE age > 18)
              limit 10;

SELECT field FROM table;
```



# Пример работы (результат)

```
(venv) [wignorbo@wignorbo sparksql_linter]$ python src/main.py src/sql/wrong_syntax.sql
wrong_syntax.sql:1:9      | E001: FROM follows immediately after SELECT [SELECT FROM;]
wrong_syntax.sql:5:15     | E001: FROM follows immediately after SELECT [SELECT from b;]
wrong_syntax.sql:5:15     | D001: Table does not exists [b]
Hint: Create table --> b
wrong_syntax.sql:7:8      | S001: Keyword should be uppercase [from]
Hint: from --> FROM
wrong_syntax.sql:7:15     | D001: Table does not exists [clients]
Hint: Create table --> clients
wrong_syntax.sql:9:8      | S002: Identifier should be in snake case [Id]
Hint: Id --> id
wrong_syntax.sql:9:16     | D001: Table does not exists [clients]
Hint: Create table --> clients
wrong_syntax.sql:9:28     | S002: Identifier should be in snake case [clientNames]
Hint: clientNames --> client_names
wrong_syntax.sql:11:5     | S001: Keyword should be uppercase [limit]
Hint: limit --> LIMIT
wrong_syntax.sql:13:19    | S001: Keyword should be uppercase [table]
Hint: table --> TABLE
```



## Ссылка на исходники

GitHub: [https://github.com/itookyourboo/sparksql\\_linter](https://github.com/itookyourboo/sparksql_linter)

