

Разработка модуля загрузки и обработки результатов автоматических тестов производительности систем хранения данных

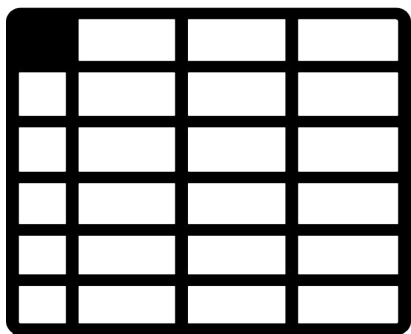
Фадеева Анастасия
группа 371
научный руководитель
ст. преп. Немешев М. Х.

Введение

- Производительность
- Тестирование
- Анализ данных



Введение



Постановка задачи

Цель: Реализовать модуль, отвечающий за загрузку и обработку результатов автоматизированных тестирований storage-систем

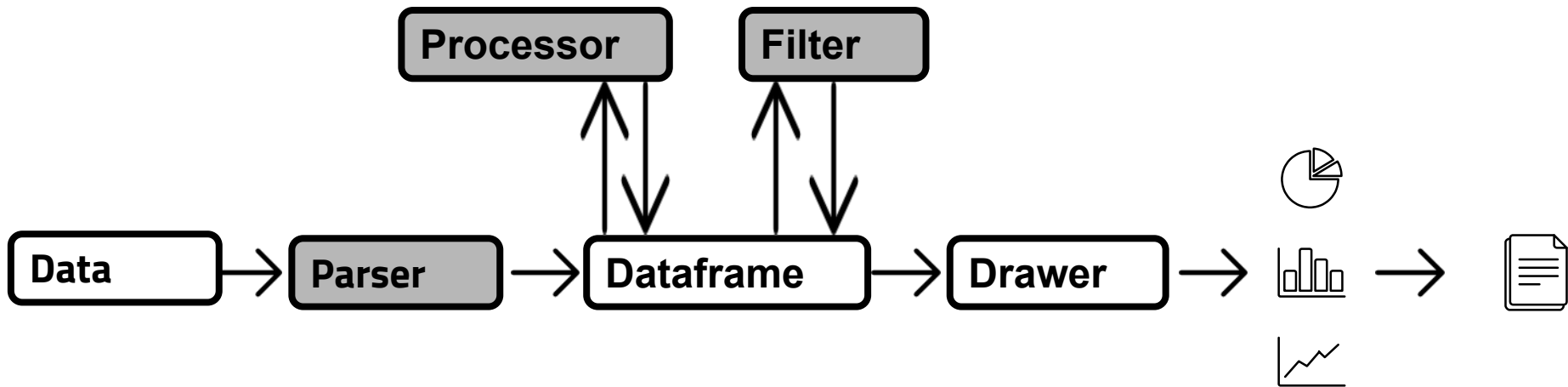
Задачи:

- проведение исследования предметной области и обзора существующих решений
- разработка расширяемого загрузчика данных различных форматов
- реализация обработчика данных, позволяющего:
 - применять различные фильтры
 - выполнять математические операции
 - группировать и трансформировать входные данные

Существующие решения

- Homebrew scripts
- Tableau
- QlikView
- Microsoft Power BI
- Excel
- Crystal Reports

Схема работы фреймворка



Результаты

Разработан **Parser** данных в различных форматах:

- **csv** - самый распространенный формат представления результатов автоматизированного тестирования
- **slg** - файлы, в которых содержится статистика по каждому элементу системы, была учтена синхронизация результатов по времени
- **compare** - файлы, в которых содержатся результаты анализа поведения системы при компрессии или дедупликации
- **vdbench** - файлы, порожденные утилитой vdbench - генератором рабочей нагрузки для измерения производительности хранилища

Также было реализовано чтение сразу нескольких файлов, названия которых соответствуют регулярному выражению, заданному пользователем.

Результаты

Реализован **Filter** со следующими функциями:

- фильтрация по регулярному выражению
- фильтрация по значению (доступны операции сравнения: le, lt, ge, gt, eq)
- убрать первые и/или последние нулевые строки

Реализован **MathProcessor**:

- max/min/mean/sum
- round
- diff
- применить свою функцию к данным

Результаты

Разработан **CommonProcessor**:

- получение датафреймов для уникальных сущностей по какой-либо колонке в виде словаря вида: {unique value: BaseDataFrame}
- группировка данных по значению в какой-либо колонке для последующего вычисления статистики
- объединение нескольких датафреймов в один