



Оптимизация загрузки новых данных в таблицы

Гурциев Ричард Зурабович
студент 3 курса физико-технического факультета

научный руководитель от ОИЯИ: Прокошин Ф.В. - кандидат физико-математических наук, старший сотрудник ЛЯП ОИЯИ

научный руководитель от ВУЗа: Тваури. И.В. - кандидат физико-математических наук, доцент, декан физико-технического факультета СОГУ



PostgreSQL - свободная объектно-реляционная система управления базами данных. Считается, что PostgreSQL хорошо подходит для работы с большими данными и в ситуациях со сложными операциями записи-чтения с одновременной валидацией данных.

Основные преимущества:

- Поддерживает функцию управления параллелизмом нескольких версий.
- Полная совместимость с SQL и ACID.
- Легкость миграции на другое решение.
- Удобство работы за счет разных индексов.
- Есть возможность сохранения произведенных процедур, за счет чего возможно расширение опций.
- Поддерживает JSON и другие функции NoSQL.
- Функции не ограничивают размер БД, даже если она исчисляется петабайтами.

Проблема

Долгое ожидание запроса

```
eventindex=# SELECT * FROM events;
```

run_number	event_number	olt_result	dsid_raw	fuid_raw
280308061	8	29721	1	3ad87450-23e6-4b72-afcb-077b5184577e
280308061	14	24218	1	3ad87450-23e6-4b72-afcb-077b5184577e
280308061	22	11794	1	3ad87450-23e6-4b72-afcb-077b5184577e
280308061	37	13903	1	3ad87450-23e6-4b72-afcb-077b5184577e
280308061	39	18546	1	3ad87450-23e6-4b72-afcb-077b5184577e
280308061	54	27200	1	3ad87450-23e6-4b72-afcb-077b5184577e
280308061	66	23844	1	3ad87450-23e6-4b72-afcb-077b5184577e
280308061	71	18992	1	3ad87450-23e6-4b72-afcb-077b5184577e
280308061	86	9718	1	3ad87450-23e6-4b72-afcb-077b5184577e
280308061	99	15973	1	3ad87450-23e6-4b72-afcb-077b5184577e
280308061	113	18546	1	3ad87450-23e6-4b72-afcb-077b5184577e
280308061	129	1296	1	3ad87450-23e6-4b72-afcb-077b5184577e
280308061	136	23743	1	3ad87450-23e6-4b72-afcb-077b5184577e
280308061	141	10936	1	3ad87450-23e6-4b72-afcb-077b5184577e
280308061	154	5041	1	3ad87450-23e6-4b72-afcb-077b5184577e
280308061	160	9718	1	3ad87450-23e6-4b72-afcb-077b5184577e
280308061	181	26502	1	3ad87450-23e6-4b72-afcb-077b5184577e
280308061	182	11376	1	3ad87450-23e6-4b72-afcb-077b5184577e
280308061	188	9144	1	3ad87450-23e6-4b72-afcb-077b5184577e
280308061	190	15356	1	3ad87450-23e6-4b72-afcb-077b5184577e
280308061	197	1881	1	6cfc4dfc-c63b-4ae2-95f3-b2e889478091
280308061	206	8011	1	6cfc4dfc-c63b-4ae2-95f3-b2e889478091
280308061	220	86	1	6cfc4dfc-c63b-4ae2-95f3-b2e889478091
280308061	228	31532	1	6cfc4dfc-c63b-4ae2-95f3-b2e889478091
280308061	235	30732	1	6cfc4dfc-c63b-4ae2-95f3-b2e889478091
280308061	252	11294	1	6cfc4dfc-c63b-4ae2-95f3-b2e889478091
280308061	257	1881	1	6cfc4dfc-c63b-4ae2-95f3-b2e889478091
280308061	264	23740	1	6cfc4dfc-c63b-4ae2-95f3-b2e889478091
280308061	282	4495	1	6cfc4dfc-c63b-4ae2-95f3-b2e889478091
280308061	294	14914	1	6cfc4dfc-c63b-4ae2-95f3-b2e889478091
280308061	315	15356	1	6cfc4dfc-c63b-4ae2-95f3-b2e889478091
280308061	321	17183	1	6cfc4dfc-c63b-4ae2-95f3-b2e889478091

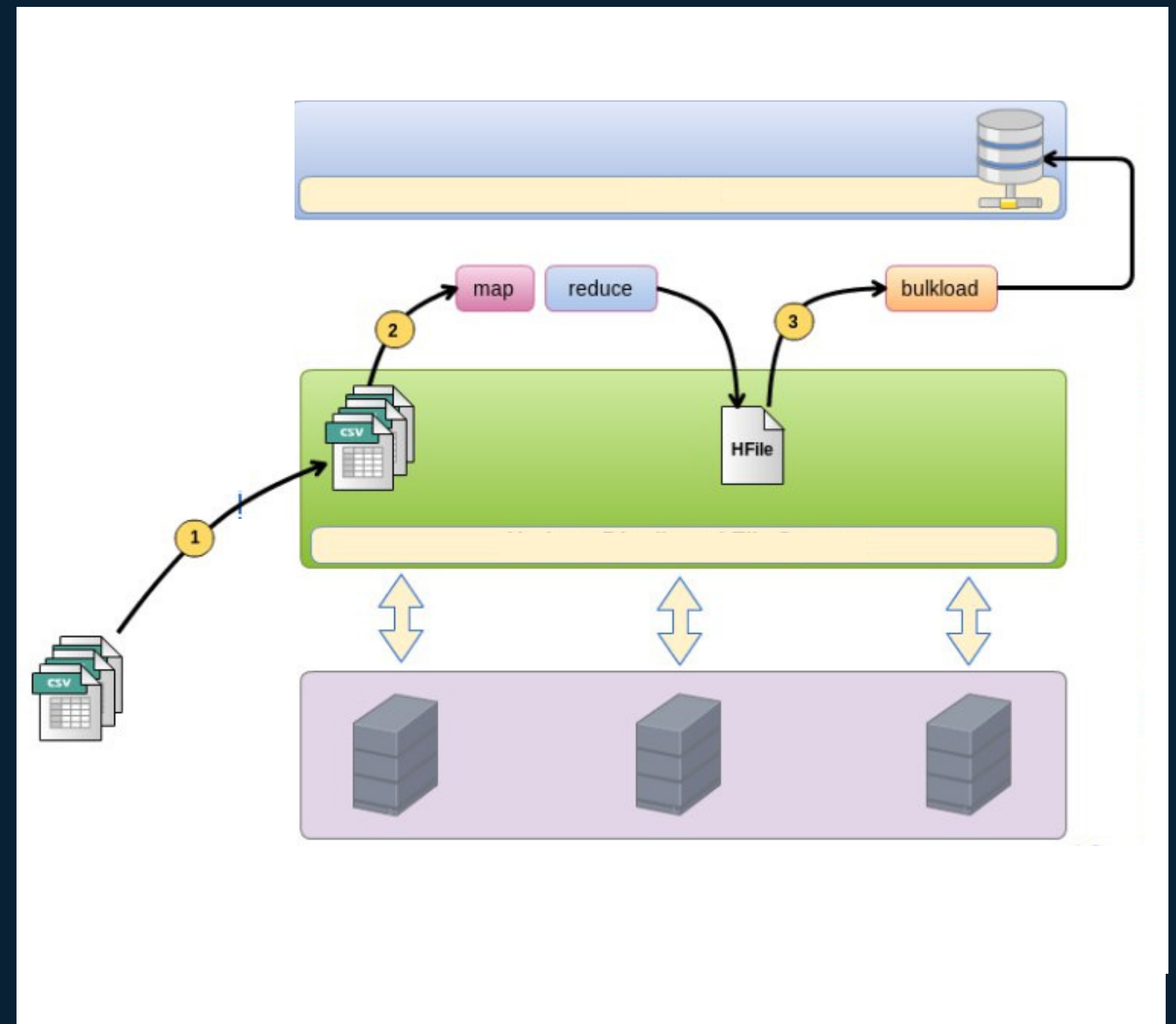
Bulk loading

Есть несколько параметров, которые следует принять во внимание, чтобы ускорить массовую загрузку огромных объемов данных с использованием PostgreSQL:

1. **COPY** вместо **INSERT**
2. Оптимизация контрольных точек
3. Незарегистрированные таблицы
4. Отключенные триггеры
5. Оптимизация порядка столбцов для массовой загрузки

Type: int8, int4, timestamptz и тд.

Type: varchar, text и тд.



asyncpg

