

**Kolegium Nauk Przyrodniczych**

**Uniwersytet Rzeszowski**

**Przedmiot:**

**Proseminarium**

**Dokumentacja projektu:**

***Porównanie Clickhouse z MS SQL, PostgreSQL***

**Wykonali:**

**Zespół projektowy: Paweł Giza, Krzysztof Pępek, Damian Pantoła, Roman Kushla**

**Prowadzący: dr inż. Piotr Grochowalski**

**Rzeszów 2021**

# Cel aplikacji

Analiza baz danych pod względem czasów wykonywania operacji na danych.

# Działanie aplikacji

* Aplikacja została zaprojektowana na dockerze.
* Aplikacja wykorzystuje do back-endu język python oraz framework django.
* Aplikacja do front-endu wykorzystuje frameworki: VueJs, Bootstrap

# Opis wykorzystanych technologii

* **Docker** - narzędzie które tworzy tzw. “wirtualne kontenery” na których można wirtualizować poszczególne elementy aplikacji.
* **Django** - Framework webowy użyty do stworzenia API do wykonywania zapytań / akcji, oparty na języku Python.
* **Vue** - Frontendowy framework webowy użyty do wyświetlania danych.
* **Clickhouse** - baza danych oparta na strukturze kolumnowej.
* **PostgreSQL** -jeden z trzech najpopularniejszych otwartych systemów zarządzania relacyjnymi bazami danych.
* **Ms SQL -** system zarządzania bazą danych, wspierany i rozpowszechniany przez korporację [Microsoft](https://pl.wikipedia.org/wiki/Microsoft).

# Uruchomienie aplikacji

* pobranie repozytorium z githuba: **https://github.com/damianpan/Proseminarium\_2021\_Damian\_Pantola\_Pawel\_Giza\_Krzysztof\_Pepek\_Roman\_Kush.git**
* pobranie danych do operacji z: <https://drive.google.com/file/d/1PcYPZWzGjK88b2n0NwKbM0lvUtgs779Q/view?usp=sharing> i wrzucenie pliku do folderu **./api**
* docker-compose build
* docker-compose up -d

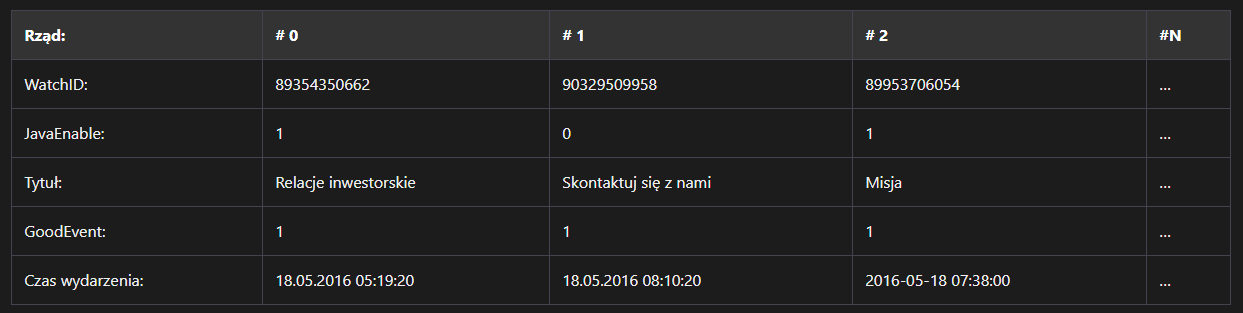
Aplikacja powinna być dostępna pod adresem: [127.0.0.1](http://127.0.0.1) w przeglądarce

# Porównanie baz danych

* **Clickhouse**

Kolumnowo zorientowany system zarządzania bazami danych. Pozwala na analizę danych w czasie rzeczywistym. Oferuje praktycznie natychmiastowe wyniki w większości wypadków, dane są przetwarzane szybciej niż tworzenie zapytania.

**Przykładowa struktura:**



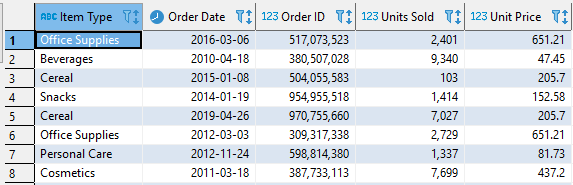
Ilość firm wykorzystujących to rozwiązanie: **22**

Dane z :<https://stackshare.io/stackups/clickhouse-vs-postgresql>

* **PostgreSQL**

Potężny, otwarty na modyfikacje dla każdego relacyjny system zarządzania bazami danych. PostgreSQL jest zaawansowany pod względem rozbudowanego podsystemu zapytań SQL, transakcji, podzapytań, triggerów, zdefiniowanych przez użytkownika typów i funkcji.

**Przykładowa struktura :**

****

Ilość firm wykorzystujących to rozwiązanie: **3626**

Dane z : <https://stackshare.io/stackups/clickhouse-vs-postgresql>

* **MsSQL**

Jest to system wykorzystywany na poziomie przedsiębiorstw, który jest bardzo popularny na serwerach WWW Windows. Jest w stanie przechowywać dowolne typy danych.

# Porównanie sterowników:

* **psycopg2**
  + duże wsparcie społeczności
  + wspiera import danych z pliku csv
  + wspiera eksport danych do pliku csv
* **pymssql**
  + średnie wsparcie społeczności
  + nie wspiera importu danych z pliku csv
  + nie wspiera eksportu danych do pliku csv
* **clickhouse\_driver**
  + słabe wsparcie społeczności
  + wspiera import danych z pliku csv
  + nie wspiera eksportu danych do pliku csv

# 

# Czas trwania wykonywania zapytań

zielony - operacja wykonana szybciej

|  |  |  |  |
| --- | --- | --- | --- |
| **Polecenie** | **Clickhouse** | **MS SQL** | **Postgre SQL** |
| Tworzenie Bazy | 2.3197 ms | 100 ms | 0.57 ms |
| Usuwanie Bazy | 20.32 ms | 31.23 ms | 13.79 ms |
| Import Danych | 105.34 sec | 2356388.501 ms  ~ 40 min | 10.31 sec |
| Export Danych | 50.64 sec | 65357.6271 ms  ~ 65 s | 13.75 sec |
| Ilość rekordów | 217.83 ms | 4332.07 ms  ~ 4.33 sec | 370.16 ms |
| Wszystkie rekordy | 4.27 sec | 68.79 ms | 4.47 sec |
| Aktualizacja wszystkich rekordów | 103.44 ms  (czas niepoprawny)  (aktualizajca usuwanie odbywa sie asynchronicznie) | 37.19 sec | 23.49 sec |
| Suma po jednej kolumnie | 203.49 ms | 283.76 ms | 580.75 ms |
| Ilość po grupowaniu | 153.08 ms | 900.48 ms | 673.58 ms |
| Pobranie danych pogrupowanych, posortowanych | 45.14 ms | 153.08 ms | 580.75 ms |

# Podgląd aplikacji

# 

# 

# Podsumowanie Projektu

* **Wnioski**

Z danych w tabelce można zauważyć, że zapytania wykonywane na bazie **Clickhouse** jeśli chodzi o pobieranie danych, obliczanie sum, grupowanie, sortowanie są szybsze, natomiast jeśli chodzi o import danych, eksport szybsza jest baza **PostgreSQL,** różnice w czasach pomiędzy różnymi systemami wynikają głównie z wirtualizacji całego środowiska. Baza **MsSQL** w porównaniu z innymi systemami jest w tyle.

**Najszybszy system z testowanych: Clickhouse**

**Najprzyjemniejszy w pracy : PostgreSQL.** Ponieważ ma bardzo duże wsparcie społeczności, Zapytania do bazy są w postaci zwykłych zapytań SQL, ( w bazie **Clickhouse** operacje do aktualizacji, usuwania danych potrzebują dodatkową składnię ), Czasów wykonywania operacji aktualizacji i usuwania nie da się sprawdzić w **Clickhousie** ponieważ są wykonywane asynchronicznie.