


Podstawy Inżynierii Danych		
Laboratorium 03		
Piotr Cofta 2024-25		
Kafka / Python		

Cel

Celem zadania laboratoryjnego jest realizacja prostych fragmentów działających oprogramowania wykorzystujących broker Kafka.

Wymagana platforma

Zadanie jest realizowane w oparciu o system Linux wraz z pakietem docker. Zależnie od indywidualnych preferencji, Linux może być zainstalowany jako maszyna wirtualna w Windows (w tym wypadku zalecane jest użycie VirtualBox), albo bezpośrednio na komputerze. Zadanie jest przygotowane dla dystrybucji Ubuntu z interfejsem graficznym, ale może być realizowane również z wykorzystaniem innych dystrybucji. Zalecane jest używanie wersji LTS.

Konieczny jest dostęp do sieci z poziomu systemu Linux. Należy zadbać o zainstalowanie pakietów docker.io. Zalecane jest ponadto zainstalowanie pakietu ntp. Może być konieczne również zainstalowanie pakietu zip oraz któregoś z edytorów.

Wymagane jest albo dodanie użytkownika do grupy docker albo do sudo. Pierwsza metoda jest zalecana.

Oprogramowanie jest pisane w języku Python, wersja 3. Dobór narzędzi programistycznych (np. PyCharm) jest pozostawiony studentowi.

Opis zadania

Instalacja plików

Wymagane pliki powinny być skopiowane na platformę Linux spod adresu podanego przez prowadzącego. Domyślnie jest to adres cofta.eu/pid3.zip. Po rozpakowaniu pliki powinny być umieszczone w jednym katalogu, tutaj domyślnie /srv/lib. W razie konieczności należy utworzyć stosowne kartoteki.

Instalacja generatora i brokera

Należy przejrzeć kod załączonego skryptu (tiny-generator.sh) oraz plików, utworzyć stosowne kartoteki (jeśli potrzebne), zmodyfikować skrypt i pliki (w miarę potrzeb) oraz uruchomić go. Poprawność instalacji należy sprawdzić poprzez komendę docker ps. Komenda powinna wykazać dwa działające kontenery.

W raporcie wykonanie zadania należy udokumentować kopią ekranu.

Sprawdzenie przepływu danych

Należy sprawdzić czy następuje przepływ danych od generatora do brokera. Wymaga to użycia narzędzi linii poleceń (kafka-console-consumer) dla monitorowania tematu.

W raporcie wykonanie zadania należy udokumentować kopią ekranu.

Modyfikacja generatora

Należy zmodyfikować pliki tak, aby generator był uruchomiony z innym identyfikatorem i inną częstotliwością generacji. Poprzez ponowne uruchomienie skryptu należy sprawdzić czy następuje przepływ danych od generatora do brokera zgodny z tymi parametrami.

Podstawy Inżynierii Danych		
Laboratorium 03		
Piotr Cofta 2024-25		
Kafka / Python		

W raporcie wykonanie zadania należy udokumentować kopią ekranu.

Wytworzenie obrazu generatora

Należy zmodyfikować / ograniczyć pliki tak, aby wytworzyć i zachować obraz generatora.

Wytworzony obraz jest zachowywany lokalnie, pod nadaną nazwą. Należy sprawdzić, czy znajduje się on w spisie dostępnych obrazów, wykorzystując polecenie docker images.

W raporcie wykonanie zadania należy udokumentować kopią ekranu.

Uruchomienie dwóch generatorów

Wykorzystując poprzednio utworzony obraz należy zmodyfikować pliki tak, aby utworzyć dwa kontenery z generatorami, każdy o innych parametrach, lecz z tego samego obrazu. Mają one różnić się identyfikatorem urządzenia oraz częstotliwością generacji danych.

W raporcie wykonanie zadania należy udokumentować kopią ekranu.

Zbudowanie filtru

Wykorzystując dostarczone pliki należy zbudować program w języku Python który będzie pobierał dane z określonego tematu i produkował komunikaty zawierające tylko jeden typ danych, np. V1 (tzn. działał jako filtr) do innego tematu. Tematy oraz typ danych powinny być podane jako parametry. Należy również zbudować skrypt, który uruchomi w kontenerach: broker Kafka, dwa generatory oraz filtr. Wyniki należy sprawdzić przy pomocy narzędzi linii polecenia.

W raporcie wykonanie zadania należy udokumentować kopią ekranu.

Warunki zaliczenia

- Obecność na laboratorium.
- Poprawne wykonanie zadania laboratoryjnego.
- Oddanie kompletnego raportu (mailem, do tygodnia po zajęciach)