


Podstawy Inżynierii Danych		
Laboratorium 02		
Piotr Cofta 2024-25		
Kafka		

Cel

Celem zadania laboratoryjnego jest realizacja prostej instalacji brokera Kafka i zaobserwowanie podstaw jego działania.

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 i docker-compose. 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.

Opis zadania

Instalacja plików

Wymagane pliki powinny być skopiowane na platformę Linux spod adresu podanego przez prowadzącego. Domyślnie jest to adres cofta.eu/pid2.zip. Po rozpakowaniu pliki powinny być umieszczone w jednym katalogu, tutaj domyślnie /srv/lib. W razie konieczności należy utworzyć stosowne kartoteki.

Instalacja jednej instancji brokera

Należy przejrzeć kod załączonego skryptu (tiny-kafka.sh) oraz plików, utworzyć stosowne kartoteki (jeśli potrzebne), zmodyfikować skrypt i pliki (w miarę potrzeb) oraz uruchomić skrypt.

Poprawność instalacji należy sprawdzić poprzez komendę docker ps.

Utworzenie tematu


Należy wykorzystać narzędzia linii poleceń do utworzenia tematu o 4 partycjach i jednej replikacji. Komenda

```
server@clusterone:/srv/lib$ docker exec -it kafka /bin/bash
[appuser@kafka ~]$
```

pozwala na dostęp do narzędzi. Utworzenie tematu wymaga komendy

```
[appuser@kafka ~]$ kafka-topics -bootstrap-server kafka:9092 --create
--topic pbs --partitions 4 --replication-factor 1
Created topic pbs.
[appuser@kafka ~]$
```

Należy sprawdzić, czy temat został utworzony

Podstawy Inżynierii Danych		
Laboratorium 02		
Piotr Cofta 2024-25		
Kafka		

```
[appuser@kafka ~]$ kafka-topics -bootstrap-server kafka:9092 --list
pbs
[appuser@kafka ~]$
```

Należy ponadto sprawdzić konfigurację tematu przy użyciu komendy

```
[appuser@kafka ~]$ kafka-topics -bootstrap-server kafka:9092 --describe --topic pbs
Topic: pbs      TopicId: 7VGprx0RQd2qis6XEmF8uQ PartitionCount: 4      ReplicationFactor: 1      Configs:
      Topic: pbs      Partition: 0      Leader: 1      Replicas: 1      Isr: 1
      Topic: pbs      Partition: 1      Leader: 1      Replicas: 1      Isr: 1
      Topic: pbs      Partition: 2      Leader: 1      Replicas: 1      Isr: 1
      Topic: pbs      Partition: 3      Leader: 1      Replicas: 1      Isr: 1
[appuser@kafka ~]$
```

W raporcie, należy to udokumentować kopią ekranu.

W raporcie należy również zinterpretować znaczenie uzyskanej informacji, w szczególności pojęć 'partition', 'leader', 'replicas' i 'isr'.

Wysyłanie i odbiór komunikatów

Należy utworzyć dwa terminale. W jednym należy uruchomić producenta, w drugim – konsumenta, poleceniami

```
[appuser@kafka ~]$ kafka-console-producer --bootstrap-server kafka:9092
--topic pbs
>
[appuser@kafka ~]$ kafka-console-consumer --bootstrap-server kafka:9092
--topic pbs
```

Należy wysłać kilka komunikatów i zaobserwować ich odbiór.

Następnie należy uruchomić konsumenta z opcją **--from-beginning**. Należy zaobserwować, czy kolejność komunikatów jest taka, jak poprzednio.

W raporcie, należy to udokumentować kopią ekranu.

Wiele procesów


Należy utworzyć min. cztery terminale. Na dwóch należy uruchomić konsumenta dla danego tematu, na pozostałych dwóch – producentów. Należy przestać komunikaty zarówno z pierwszego jak i z drugiego producenta. Należy zaobserwować pojawianie się nowych komunikatów, w tym ich kolejność.

W raporcie należy to udokumentować widokiem ekranu z terminalami.

W raporcie należy wyjaśnić zaobserwowane zachowania.

Grupa konsumentów

Należy utworzyć terminal z producentem dla danego tematu. Należy również utworzyć terminal z jednym konsumentem w danej grupie, np. poleceniem.

Podstawy Inżynierii Danych		
Laboratorium 02		
Piotr Cofta 2024-25		
Kafka		

```
[appuser@kafka ~]$ kafka-console-consumer --bootstrap-server kafka:9092
--topic pbs --group alpha
```

Należy przesłać kilka komunikatów i zaobserwować ich odbiór.

Teraz należy utworzyć kolejny terminal z drugim konsumentem z tej samej grupy. Należy przesłać kilka komunikatów i zaobserwować ich odbiór. Należy zaobserwować które komunikaty trafiają do których konsumentów.

W raporcie należy udokumentować widokiem ekranu z terminalami.

W raporcie należy wyjaśnić zaobserwowane zachowania.

Warunki zaliczenia

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