

Kaflog – wydajność

W celu zbadanie wydajności systemu, przeprowadzono następujące testy:

1. Czas pobierania logów z brokera i zapisywania ich do HDFS, w zależności od ilości logów i ilości partycji (parametr brokera - *num.partitions*), do których zapisywane były logi
2. Czas generowania widoków specyficznych dla systemu przy użyciu **Apache Hive**.

Testy przeprowadzane były na wirtualnym środowisku zestawionym przy użyciu narzędzia **Vagrant**, na maszynie fizycznej o następujących parametrach:

Lenovo Y580

CPU: Intel i5-3230M

RAM: 16GB

HDD: 1TB, 5600 obr/s

Środowisko składało się z następujących maszyn:

cloudera-master: 5GB RAM

- instalacja klastra **Cloudera**
- umiejscowienie broker'a

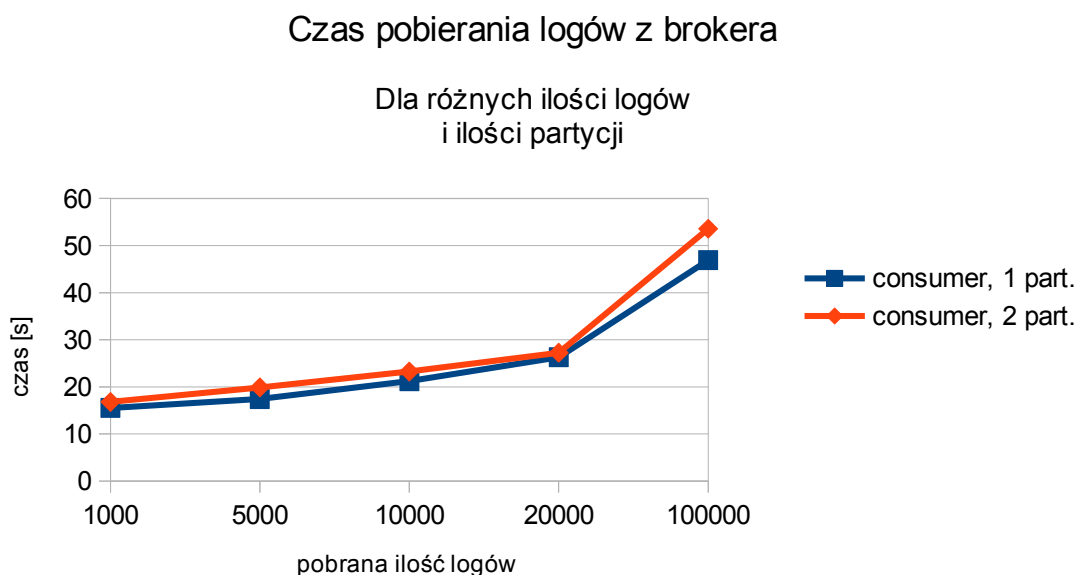
cloudera-slave1: 2GB RAM

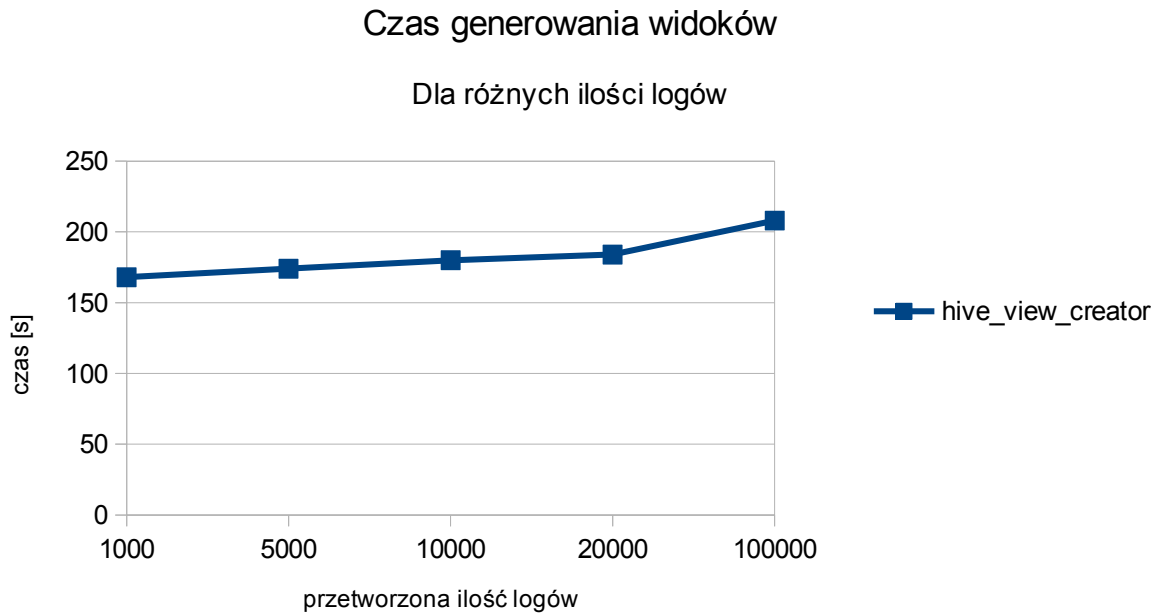
- HDFS data node
- dodatkowa instancja **Zookeeper'a**

kafka-node1: 2GB RAM

- producent logów

Poniższe wykresy prezentują wyniki otrzymane w wymienionych testach:





Propozycje optymalizacji środowiska

- rozproszenie systemu na fizycznej topologii
- umieszczenie kilku instancji brokerów na osobnych maszynach
- rozproszenie HDFS'a na co najmniej 3 maszyny fizyczne
- dodanie kolejnych instancji Zookepera
- regulacja parametrów brokera:
 - ilość partycji
 - ilość wątków obsługujących połączenia sieciowe i operacje IO