

Flume

Instalacja flume:

1. Wejdź na interfejs webowy cludera managera <http://vm-cluster-node1:7180>
2. Kliknij strzałkę w dół po prawej stronie napisu 'Cluster 1', a następnie 'Add Service'
3. Wybierz z listy 'Flume' i kliknij continue
4. kliknij na 'Select hosts'. Zaznacz checkbox przy 'vm-cluster-node2', a następnie kliknij 'OK' i 'Continue'.
5. kliknij 'Finish'
6. kliknij na nowo powstały serwis 'Flume', a następnie wejdź w zakładkę 'Configuration'
7. Przekopiuj konfigurację z pliku 'flume.conf' do pola 'Configuration File'
8. Dodaj do 'Plugin directories' wartość '/home/vagrant/flume.plugins.d'
9. Kliknij 'save changes'
10. Utwórz na vm-cluster-node2 katalog '/home/vagrant/flume.plugins.d/interceptors/lib' i wrzuć do niego zbudowanego jara z interceptorami.
11. Kliknij u góry na 'Actions'
12. Kliknij akcję 'Start'
13. Wejdź na <https://github.com/jpodeszwik/Fake-Apache-Log-Generator> i sklonuj repozytorium
14. Postępuj zgodnie z instrukcją instalacji w README
15. uruchom generator spreparowanych logów apache'a za pomocą komendy './run.sh vm-cluster-node2 9999'
16. wylistuj katalog na hdfsie /user/flume/events i zobacz czy pojawiają się logi apache'a

Zadania:

1. Napisz interceptor, który odkoduje datę z liniiki loga i ustawi odpowiednią datę w nagłówku. Skorzystaj z nagłówka 'timestamp' i umieść w nim unixowy timestamp wyznaczony na podstawie daty odkodowanej z eventu. Eventy z przedwczorajszą datą wygeneruj za pomocą polecenia `'./run2.sh vm-cluster-node2 9999'`
2. Skorzystaj z możliwości multiplexacji, żeby odfiltrować błędne linijki loga. Przekieruj błędne linijki do innego katalogu. Błędne eventy wygeneruj za pomocą komendy `'./run3.sh vm-cluster-node2 9999'`

<https://flume.apache.org/FlumeUserGuide.html#multiplexing-the-flow>