Konspekt zajęć z technologii Splunk

Rafał Grzeszczuk

1. Wszystkie polecenia wykonuj jako użytkownik splunk (chyba, że prowadzący powie inaczej)
2. Dokumentuj wykonywane polecenia i ich skomentowane wyniki
3. Na koniec laboratorium wyślij plik pdf z poleceniami (pkt.2)
4. Ręczna modyfikacja plików konfiguracyjnych wymaga ponownego uruchomienia serwera.
5. Na końcu instrukcji znajdują się przydatne linki, m.in. do dokumentacji oraz tutoriali.

Laboratorium 1

1. Instalacja i uruchomienie instancji standalone na przygotowanej maszynie wirtualnej.

a) Rozpakuj przygotowane archiwum do katalogu /opt/splunk. Następnie przejdź do katalogu /opt/splunk/bin i wywołaj polecenie *splunk start* - jeśli to konieczne, zaakceptuj umowę licencyjną. Ustaw domyślny login administratora na *admin*, a hasło na *admin123*. Sprawdź działanie serwera wchodząc na adres http://<ip>:8000/[1] w dowolnej przeglądarce internetowej.

b) W przypadku problemów z uruchomieniem serwera sprawdź, czy użytkownik splunk jest właścicielem swojego katalogu oraz czy na portach TCP 8000 i 8089 nie nasłuchują inne usługi (np. poprzednie instancje serwera)

c) Korzystając z dokumentacji pliku server.conf sprawdź w jaki sposób ustawić nazwę serwera. Ustaw ją na *INazwisko*. Uwaga: nie modyfikuj pliku w podkatalogu *default*, tylko *local*

d) Zapoznaj się z poleceniem btool [1] - wybierz dowolną własność z pliku server.conf w podkatalogu *default*, a następnie nadpisz ją w pliku server.conf w podkatalogu . Korzystając z polecenia btool sprawdź, który plik ma wyższy priorytet.

local

2. Indeksowanie danych:

a) Stwórz nowy indeks o nazwie *index_INazwisko* - możesz to zrobić edytując plik indexes.conf lub wydając polecenie:
splunk add index <nazwa> [2]

b) Zaindeksuj dane z przygotowanego pliku *lab-log.zip*. Możesz skorzystać z webgui lub wydając polecenie: *splunk add oneshot example.log* [3]. Plik możesz pobrać za pomocą polecenia wget z lokalizacji podanej przez prowadzącego.
Zapoznaj się z organizacją danych na dysku [4]. Co możesz powiedzieć o wewnętrznej strukturze plików? A o ich organizacji na dysku?

c) Powtórz czynności z punktów 2a-2c, ale tym razem skonfiguruj *monitorowanie* pliku w trybie ciągłym. Możesz skorzystać np. z logów systemowych maszyny wirtualnej, na której pracujesz.

3. Wyszukiwanie zaindeksowanych danych

a) Zapoznaj się ze składnią zapytań w języku SPL [5, 6]

b) Korzystając z indeksów stworzonych w punkcie 2, ułóż zapytania wyświetlające:
   i. wszystkie zdarzenia w przykładowym logu dotyczące uwierzytelniania
   ii. wszystkie zdarzenia dotyczące serwera http
   iii. średni rozmiar danych zwróconych przez zapytanie w ostatnim miesiącu
   iv. wizualizację średniego rozmiaru danych dla każdej kategorii (categoryId) z agregacją dobową
   v. maksymalny rozmiar transferu danych dla każdego klienta (identyfikowanego przez IP)
   vi. listę user agentów dla każdego adresu IP
   vii. dla każdej podstrony – ostatni kod http jaki wystąpił oraz poprzedni kod http jaki wystąpił wraz z czasami ich wystąpienia (podpowiedź: streamstats)

[1] Adres IP możesz sprawdzić wydając polecenie *ip addr*

c) Przygotuj zapytanie (...) oraz odpowiednią dla niego wizualizację wyników. Zapisz wizualizację jako dashboard z możliwością wyboru zakresu czasu. Udostępnij dashboard w aplikacji Search&Reporting - dzięki temu każdy użytkownik z dostępem do tej aplikacji będzie mógł skorzystać z dashboardu.
d) Zbadaj co najbardziej wpływa na czas przetwarzania zapytań. Skomentuj rezultaty, odnosząc się do wyników zadania 2c.

Laboratorium 2

4. Modele danych

a) W oparciu o wewnętrzne logi audytowe (*index=_audit*), przygotuj model danych[7] zawierający co najmniej: czas wystąpienia zdarzenia, użytkownika, którego zdarzenie dotyczy, wykonana akcja i jej szczegóły (pole *info*)
b) Policz statystyki z ostatniego tygodnia dotyczące:
   i. ilości wystąpień poszczególnych akcji per użytkownik
   ii. najczęściej pojawiającej się akcji per użytkownik

   korzystając z wyszukiwania w indeksie oraz polecenia *stats*, a następnie bezpośrednio z datamodelu (poleceniem | *tstats*) [8]. Porównaj czasy wykonania.
c) Włącz akceleracją utworzonego w podpunkcie a) datamodelu i powtórz eksperyment z podpunktu b). *Uwaga: zbudowanie/przebudowanie metadanych służących do akceleracji indeksu może zająć kilka minut (w dużych środowiskach produkcyjnych często jest to >24h). Możesz to zweryfikować w zakładce Settings -> Data Models w panelu administracyjnym webgui lub sprawdzając odpowiedni plik .tsidx.*

5. Polityka retencji

a) Zapoznaj się z etapami życia bucketów [9]. Korzystając z ustawień w pliku *indexes.conf* zapewnij następujące zachowanie indeksu:
   i. Maksymalny rozmiar bucketu otwartego do zapisu – 2MB
   ii. Maksymalna ilość bucketów *warm* – 3
   iii. Czas do zamrożenia bucketu – 15 minut
   iv. Lokalizacja zamrożonych bucketów - /tmp/INazwisko/frozen
b) Dokonaj pomiaru wydajności wyszukiwania w poszczególnych bucketach.
c) Czy w sytuacji posiadania znacznych zasobów dyskowych uzasadnione jest trzymanie wszystkich danych w bucketach *hot*? Odpowiedź uzasadnij.
d) Korzystając z komendy *splunk rebuild* oraz instrukcji [10] przywróć do wyszukiwania dowolny zamrożony bucket. Ile trwa operacja przywracania? Jaka jest różnica w rozmiarze między bucketem *frozen* a przywróconym? Jak wydajne jest wyszukiwanie w przywróconym buckecie?

6. Forwarding danych - universal forwarder

a) Pobierz oprogramowanie Splunk Universal Forwarder i zainstaluj je na maszynie wirtualnej z systemem operacyjnym Windows [11].
b) Skonfiguruj indeksowanie logów audytowych systemu Windows
c) Zweryfikuj poprawność zaindeksowanych danych oraz automatycznej ekstrakcji pól. Czy występują różnice między czasem wystąpienia zdarzenia a czasem indeksacji?
d) Zasymuluj awarię sieci pomiędzy forwarderem a indekserem – po 5 minutach „napraw” połączenie. Powtórz czynności z punktu c – jakie wyciągniesz wnioski?

Przydatne linki

1. https://docs.splunk.com/Documentation/Splunk/7.2.1/Troubleshooting/Usebtooltotroubleshootconfigurations
2. https://docs.splunk.com/Documentation/Splunk/7.2.3/Admin/CLIadmincommands
3. https://docs.splunk.com/Documentation/Splunk/7.2.3/Data/MonitorfilesanddirectoriesusingtheCLI
4. https://docs.splunk.com/Documentation/Splunk/7.2.2/Indexer/HowSplunkstoresindexes
5. https://docs.splunk.com/Documentation/SplunkCloud/7.1.3/SearchTutorial/Startsearching
6. https://www.youtube.com/watch?v=xtyH_6iMxwA
7. https://docs.splunk.com/Documentation/Splunk/7.2.3/Knowledge/Managedatamodels#Create_a_new_data_model
8. https://docs.splunk.com/Documentation/Splunk/7.2.3/SearchReference/Tstats
9. https://docs.splunk.com/Documentation/Splunk/7.2.3/Indexer/Setaretirementandarchivingpolicy
10. https://docs.splunk.com/Documentation/Splunk/latest/Indexer/Restorearchiveddata
11. https://docs.splunk.com/Documentation/Forwarder/7.2.3/Forwarder/HowtoforwarddatatoSplunkEnterprise

Dla zainteresowanych

- https://answers.splunk.com/answers/494928/what-exactly-is-a-tsdix-file.html