Projecte final de curs

Autor

- Kevin Ramos Lopez
- isx47752902
- ASX-2
- Curs 2015/2016
- Project-GitHubRepo

Descripció

Aquest es el treball elaborat com a projecte de final de curs de ASIX. El tema general d'aquest any es **tecnologies actuals per la gestió de logs**. I el tema concret escollit per aquest projecte es **Centralització de logs JSON en BBDD i post-processat**.

Breu d'escripció del projecte:

Actualment existeixen BBDD pensades per a emmagatzemar documents en format JSON. Això permet que la sortida en format JSON del journal de diverses màquines s'emmagatzemi de manera centralitzada per a un posterior processat. En aquest àmbit existeixen bases de dades com Elasticsearch o Rethinkdb. Per al post-processat existeix l'entorn ELK (Elasticsearch + Kibana) que està pensat per anàlisi de sèries temporals de dades com són els logs.

Serveis utilitzats:

- Logstash
- Elasticsearch
- Kibana

ELK

La utilització d'aquets serveis per processar les dades s'està fent cada cop més popular, la combinació dels tres serveis mencionats anteriorment s'anomena (ELK). Proporcionen informació processable a temps real de casi cualsevol tipus de font de dades estructurades o no estructurades. Milers de organitzacions s'han sumat la utilització d'aquest servei:

- Netflix
- Stack Overflow
- Linkedin
- Fujitsu (Open Stack Cloud)

I moltes mes...

Logstash

Logstash Official Documentation

Definició

Logstash es un motor de recopilació de dades open source. El cual pot unificar diferents entrades de dades i normalitzar aquestes en el destí indicat.

Funcionament

El funcionamnt de logstash es basa en un fitxer de configuració, en el cual s'indiquen dos parts obligatories "Input" i "Output" i una opcional "Filter". A partir de la versió 2.3 el proces de filtratge i el de sortida es realitzen en la mateixa etapa, cosa que millora el rendiment.

La ubicació del fitxer de configuració es `/etc/logstash/conf.d/`

Al iniciar una instancia logstah, s'indica amb -f el fitxer de configuració que definirà la instancia.

```
logstash -f /etc/logstash/conf.d/name.conf
```

Exemple bàsic del config file

```
input {

    file {

        path => "/var/log/proba/apache-logs.log"

        start_position => "beginning"
```

```
        type => "apache"
    }

filter {

        grok {

            match => { "message" => "%{COMBINEDAPACHELOG}"}

        }

        geoip {

            source => "clientip"

        }

}

output {

        elasticsearch {

            hosts => ["hostElast:9200"]

            index => "logstash-%{+YYYY.MM.dd}"

        }

}
```

- Input: S'indica quina serà l'entrada de logs a procesar, en aquest cas un fitxer local.
- Filter: Defineix els filtres que s'aplicaràn als logs d'entrada.
  - grok: Analitza I estructura el text, Es la millor eina de logstash per convertir els logs no estructurats en algo estructurat I consultable.
  - geoip: Afegeix informacio de la localització de la IP. (A mes mostra grafiques en Kibana).
- Output: Serà la sortida d'aquets logs, en el nostre cas, l'output serà el servidor elasticsearch, on es guardaran en format JSON.

Exemple avançat del config file logstash-pipeline.conf

Aquest ja es un fitxer mes complet:

- Input: Veiem que hi han 3 entrades diferents de logs, per identificarles l'hi assignem un `type` a cada una.
- Filter: Apliquem els filtres nomes als logs de `type` apache
- Output: Segons el `type` creem els index que es guardaràn a la BBDD d'elasticsearch.

Elasticsearch

Elasticsearch Official Documentation

Definició

Es un motor de busqueda open source basat en apache. Ens permet enmagatzemar, buscar i analitzar grans cantitats de dades a temps casi real. Elasticsearch funciona en un clúster, en el cual es repliquen les dades entre els diferents nodes.

El servidor elasticsearch pot tenir tres estats:

- Red: Hi ha fragments no assignats en el clúster.
- Yellow: Tots els fragments estàn assignats, pero no hi han replicas d'alguns d'aquets.
- Green: Tots els fragments están assignats i amb les seves respectives repliques.

Per funcionar, el sistema ha de disposar de una versió recent de java.

El port asociat a aquest servei es el 9200.

Conceptes bàsics

- Clúster: Grup d'un o mes servers, el nom del cluster que es crea per defecte es "elasticsearch".

- Node: Un node que forma part del cluster, enmagatzema les dades i repliques, i participa en la indexació d'aquestes. Cada node s'identifica per un nom.

- Index: Un conjunt de documents de característiques similars. Cada index s'identifica per un nom, normalment seguit de la data de creació d'aquest.

- Type: Es una caegoría d'un index. Ja que d'ins d'un mateix index poden haber documents de diferents tipus.

- Document: Unitat bàsica d'informació que pot ser indexada. Es guarda en format JSON. Cada document te que estar assignat a un type d'ins de cada index.

- Shards i repliques: Un index pot contenir molta informació, es per aixo que aquets es divideixen en fragments, i es realitzen repliques de cada fragment, que es guardaràn en diferents nodes.

Funcionament

Per interactuar amb elasticsearch es pot utilitzar cualsevol eina que ens permeti fer crides HTTP/REST. Com per exemple curl:

- Rebre informació bàsica del clúster (nom, número de nodes, status...):

```
curl 'localhost:9200/_cat/health?v'
```

- Informació bàsica dels nodes que composen el cluster:

```
curl 'localhost:9200/_cat/nodes?v
```

- Llistat dels index que conté el clúster:

```
curl 'localhost:9200/_cat/indices?v'
```

Tambe es poden crear, modificar, borrar i consultar index a traves de curl's, pero no ens centrarem en aquesta part, ja que per al postprocesat de les dades utilitzarem kibana. Pero aquí teniu un docuemnt amb exemples de com fer-ho:

Interactuant amb elasticsearch mitjançant curl's

Podeu trobar mes informació sobre interactuar amb elasticsearch a la documentació oficial indicada a la capçalera del fitxer.

Una manera mes comode de treballar amb elasticsearch es instalant algún plugin que proporcioni una interficie gràfica, com per exemple el "head":

```
elasticsearch/bin/plugin install mobz/elasticsearch-head
```

Accedirem a traves de la url: `http://localhost:9200/_plugin/head/`

Kibana

Kibana Official Documentation

Definició

Es una plataforma d'ànalisis i visualització de dades. S'utilitza per interactuar amb les dades enmagatzemades d'elasticsearch.

Pot realizar facilment analisis avançats de dades i visualitzar aquestes a traves de gràfics i mapes.

D'aquesta manera utilitzant kibana es pot treballar amb grans volums de dades de manera fàcil i ràpida. I al estar disenyat per treballar amb elasticsearch, no presenta ningun tipus de dificultat ni requereix ningún tipus de configuració extra. Es tant facil com instalar-ho i ja es podràn visualitzar les dades del servidor elasticsearch indicat.

S'accedeix a kibana a traves del port 5601.

Funcionament

kibana disposa d'un fitxer de configuració .yml, amb l'edició d'aquest es pot configurar:

- El port que utilitzarà kibana

- El servidor elasticsearch que conté les dades que ha de post-procesar.
- Mecanismes de seguretat per comunicarse amb el server d'elastic. Poden ser passwords i usuaris, o certificats i claus ssl.
- Configuracions de timeout.
- etc...

Exemple de fitxer de configuració:

kibana.yml

Accedirem a la interficie de kibana a través de la url:

`http://localhost:5601`

Kibana es una eina completament gràfica, i la millor manera de familiaritzar-se es a través d'aquets video-turorials oficials:

Primers pasos amb kibana

Exemples de visualitzacions

Kibana ofereix un metode de lo mes sencill pero poder exportar les nostres visualitzacions, i implementarles en una pàgina web per exemple.

Es tan sencill com crear un dashboard, a la toolbar veurem una icona on posa "shared", i d'allà podrem treure un sencill link, el cual al introduirlo en un fitxer html o markdown, s'estarà visualitzant el nostre dashboard.