

Laboratorio de Sistemas Operativos y Redes

**Open Source Search Server**



## **Instalación y configuración**

Profesor:

**José Luis Di Biase**

Integrantes:

**Martín Kopecek**

**Leonel Martos**

**Camila Sosa**

Fecha: 12/7/2018

# 1. ¿Qué es Sphinx?

Sphinx es un motor de búsqueda de código abierto que provee la funcionalidad de búsqueda de texto a aplicaciones cliente. Es conocido por realizar búsquedas eficientes sobre grandes cantidades de datos. Fue especialmente diseñado para integrarse bien con las bases de datos SQL almacenando los datos, y para accederlos fácilmente mediante lenguajes de scripting. Funciona indexando contenidos de bases de datos.

El desarrollo de Sphinx comenzó en 2001 por su autor Andrew Aksyonoff, y continúa en desarrollo hasta el día de hoy por la empresa homónima, con su última versión lanzada en marzo de 2018.

En nuestro trabajo, vamos a utilizar SphinxQL, con el que nuestra aplicación (MySQL en nuestro caso) consulte Sphinx usando la librería estándar del cliente de MySQL y la sintaxis de consultas del mismo.

## 2. Requisitos para la instalación

Vamos a realizar la instalación en una distribución de Linux, en particular Ubuntu 16.04

Necesitamos:

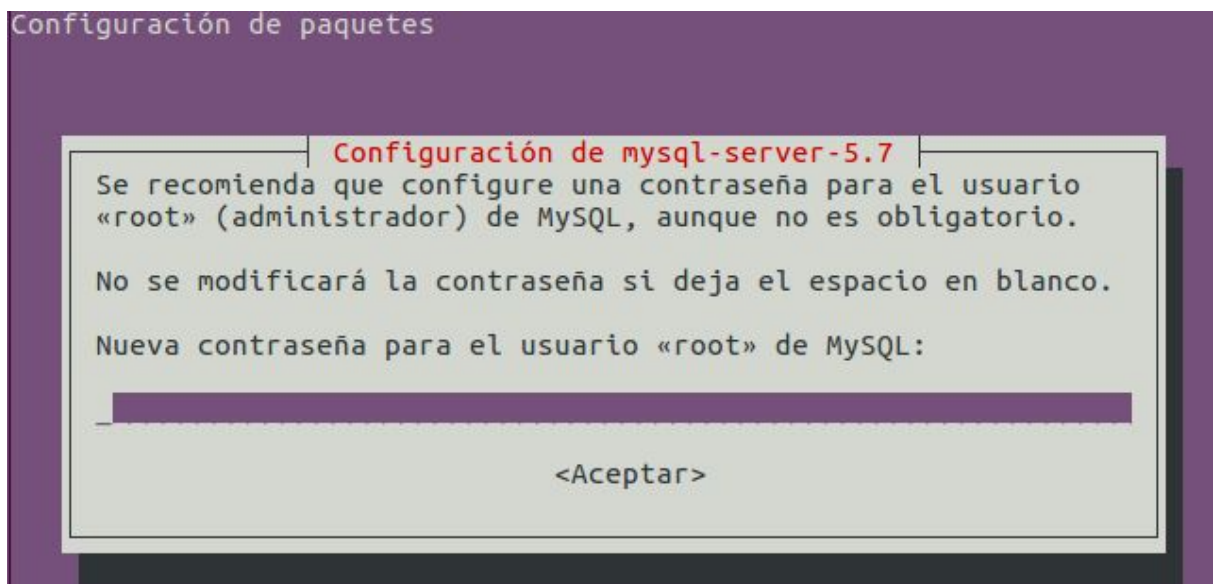
- Un usuario sudo.
- Tener instalado MySQL (Para probar su funcionamiento).

### 2.1. Instalando MySQL

Podemos instalar MySQL con el siguiente comando:

```
$ sudo apt-get install mysql-server
```

Aceptamos la descarga de los paquetes y al terminar, se abrirá una ventana para que configuremos la contraseña de *root*, si así lo deseamos:



Escribimos la contraseña y la volvemos a escribir para confirmar. Luego de esto se van a desempaquetar los paquetes y finalizará su instalación.

## 3. Instalación de Sphinx

### 3.1 Instalando Sphinx

Para instalarlo, simplemente usamos el siguiente comando:

```
$ sudo apt-get install sphinxsearch
```

### 3.2 Creando una base de datos de prueba

Primero nos logueamos en MySQL:

```
camila@camila-VirtualBox:~$ mysql -u root -p  
Enter password: █
```

Nuestro prompt cambiará a *mysql/*>.

Ahora creamos nuestra base de datos de prueba:

```
mysql> CREATE DATABASE test;
```

Importamos el archivo de prueba SQL:

```
mysql> SOURCE /etc/sphinxsearch/example.sql
```

Y por último salimos del shell de MySQL:

```
mysql> quit
```

## 4. Configurando Sphinx

Necesitamos crear un archivo llamado *sphinx.conf* en el directorio */etc/sphinxsearch* :

```
$ sudo nano /etc/sphinxsearch/sphinx.conf
```

El archivo consiste de tres bloques esenciales para correr:

### source:

```
source src1
{
  type          = mysql

  sql_host      = localhost
  sql_user      = root
  sql_pass      = 1234
  sql_db        = test
  sql_port      = 3306

  sql_query     = SELECT * FROM documents
  sql_attr_uint = group_id
  sql_attr_timestamp = date_added
}
```

- **sql\_host**: El host en dónde se encuentra la base de datos, puede ser localhost o una dirección IP.
- **sql\_user**: El usuario con el que me quiero conectar a la base de datos.
- **sql\_pass**: La contraseña de dicho usuario.
- **sql\_db**: El nombre de la base de datos.
- **sql\_port**: El puerto por el cual me conecto.

- **sql\_query**: La query con la que vuelca los datos de la base de datos a el index.

### index:

```
index test1
{
  source = src1
  path = /var/lib/sphinxsearch/data/test1
  docinfo = extern
}
```

Contiene el source y la ruta en donde almacena los datos.

### searchd:

```
searchd
{
  listen = 9306:mysql41
  log = /var/log/sphinxsearch/searchd.log
  query_log = /var/log/sphinxsearch/query.log
  read_timeout = 5
  max_children = 30
  pid_file = /var/run/sphinxsearch/searchd.pid
  seamless_rotate = 1
  preopen_indexes = 1
  unlink_old = 1
  binlog_path = /var/lib/sphinxsearch/data
}
```

Contiene el puerto en donde el daemon de SphinxQL está escuchando y otras variables para correrlo.

## 5. Manejando el Index

Un Index es una estructura de datos que permite recuperar las filas de una tabla de forma más rápida. Sphinx construye dichas estructuras a partir de los datos.

Primero, agregamos data al Index usando la configuración que creamos previamente:

```
$ sudo indexer --all
```

**PROBLEMA** : al hacer varios cambios en el archivo `sphinxsearch.conf` el comando `sudo indexer --all` puede lanzar el siguiente error:

**FATAL: failed to lock /var/lib/sphinxsearch/data/test1.spl: Resource temporarily unavailable, will not index. Try --rotate option.**

Esto se soluciona borrando el archivo `test1.spl` ubicado en `/var/lib/sphinxsearch/data/` con el comando `rm`

**indexer** una de las herramientas principales de Sphinx, se encarga de manejar la data (indexes) en los cuales realizaremos nuestras búsquedas.

El parámetro `--all` actualiza todos los indexes.

Deberíamos tener el siguiente output:

```
Sphinx 2.2.9-id64-release (rel22-r5006)
Copyright (c) 2001-2015, Andrew Aksyonoff
Copyright (c) 2008-2015, Sphinx Technologies Inc (http://sphinxsearch.com)

using config file '/etc/sphinxsearch/sphinx.conf'...
indexing index 'test1'...
collected 4 docs, 0.0 MB
sorted 0.0 Mhits, 100.0% done
total 4 docs, 193 bytes
total 0.075 sec, 2556 bytes/sec, 52.97 docs/sec
total 4 reads, 0.000 sec, 0.1 kb/call avg, 0.0 msec/call avg
total 12 writes, 0.000 sec, 0.1 kb/call avg, 0.0 msec/call avg
```

## 6. Comenzando el servicio

Por defecto, el daemon de Sphinx está apagado. Lo activaremos cambiando la línea *START=no* a *START=yes* en el archivo */etc/default/sphinxsearch* :

```
$ sudo sed -i 's/START=no/START=yes/g' /etc/default/sphinxsearch
```

Usamos *systemctl* para reiniciar el daemon de Sphinx:

```
$ sudo systemctl restart sphinxsearch.service
```

Chequeamos que esté corriendo correctamente:

```
$ sudo systemctl status sphinxsearch.service
```

Deberíamos obtener algo como esto:

```
● sphinxsearch.service - LSB: Fast standalone full-text SQL search engine
   Loaded: loaded (/etc/init.d/sphinxsearch; bad; vendor preset: enabled)
   Active: active (exited) since mié 2018-07-04 01:05:44 -03; 27s ago
     Docs: man:systemd-sysv-generator(8)
```

Así nuestro daemon de Sphinx está corriendo correctamente.



## 7. Probando Sphinx localmente

Iniciamos MySQL conectándonos al daemon de Sphinx con el siguiente comando:

```
$ mysql -h0 -P9306 --prompt='sphinxQL> '
```

**ERROR** : al escribir esto en consola es posible recibir el error **ERROR 2003 (HY000): Can't connect to MySQL server on '0' (111)** para solucionar esto reiniciamos el servicio de sphinxsearch escribiendo en consola :  
`systemctl restart sphinxsearch.service`

Ya conectados, nuestro prompt cambiará a *sphinxQL>* . Hacemos el cambio explícito ya que en caso contrario el prompt dirá *mysql>* y queremos tener presente que estamos trabajando con Sphinx.

Con el comando SHOW TABLES podemos ver el index que contiene la información de la base de datos test.

```
sphinxQL> SHOW TABLES;
+-----+-----+
| Index | Type |
+-----+-----+
| test1 | local |
+-----+-----+
1 row in set (0,00 sec)
```

Procedemos a hacer una consulta:

```
sphinxQL> SELECT * FROM test1 WHERE(id=1); SHOW META;
+-----+-----+-----+
| id    | group_id | date_added |
+-----+-----+-----+
| 1     | 1        | 1530676738 |
+-----+-----+-----+
1 row in set (0,00 sec)

+-----+-----+
| Variable_name | Value |
+-----+-----+
| total         | 1     |
| total_found   | 1     |
| time          | 0.000 |
+-----+-----+
3 rows in set (0,00 sec)
```

Nos devuelve la tupla cuyo id es igual a 1. Debajo del resultado de la consulta, nos muestra información acerca del resultado.

