

理论篇

一、Sphinx[sfɪŋks]优势及概述

- a) 优势：站内全文搜索，优于 MySQL 搜索引擎

MySQL-->Sphinx(生成索引文件) <--查询---返回结果-->Sphinx 提供 phpAPI

- b) 概述

一个基于 SQL 的全文检索引擎 ,它提供了比数据库更专业的搜索功能 ,并且 sphinx 特别为一些脚本语言设计了搜索的 API 接口 ,如 PHP , Python 等 ,同时为 MySQL 也设计了一个存储引擎插件。

Sphinx -----> MySQL

i. MySQL 采用创建索引

1. mysql> SELECT count(1) FROM cdr_file // 1703122 条记录
2. 查询数据表索引 SHOW INDEX FROM|IN `table_name`;
3. 删除索引 DROP index `index_name` ON `table_name`
ALTER TABLE `table_name` DROP INDEX|UNIQUE|PRIMARY KEY `index_name`
4. 创建索引
ALTER TABLE `table_name` ADD INDEX|UNIQUE|PRIMARY KEY| `index_name`
(column list);
mysql> ALTER table cdr_file ADD INDEX index_transactionid(transactionid);
Query OK, 0 rows affected (56.48 sec)

ii. 创建索引模糊查询

```
mysql> select count(1) from cdr_file where transactionid like '81%';
+-----+
| count(1)|
+-----+
| 21791   |
+-----+
1 row in set (9.82 sec)
```

iii. 创建索引模糊查询

```
select count(1) from cdr_file where transactionid like '81%';
+-----+
| count(1) |
+-----+
| 21791 |
+-----+
1 row in set (0.17 sec)
```

二、Sphinx 应用场景

a) 网站中的搜索（站内搜索）

- i. 比如在优酷网站内搜索时，秒查出来列表
- ii. 传统方式：PHP ---> MySQL
- iii. 优化方式：1)MySQL <---生成索引文件--->Sphinx 2)PHP ----->Sphinx

b) 系统后台中的搜索

三、Sphinx 进阶

a) 为什么需要 Sphinx

- i. 做搜索时当数据量大单纯的 MySQL 搜索比较慢(使用模糊或者联合查询如果表进行了分表更慢)
- ii. 搜索是中文分词（即搜索词都会被包含搜索出来）
- iii. Sphinx 查询快
 1. 高速的建立索引（在当代 CPU 中。峰值性能达 10M/s）
 2. 高性能的搜索（在 2~4G 的文本数据上，平均每次检索时间小于 0.1 秒）
 3. 可处理海量数据

b) 如何获取 sphinx

- i. <http://sphinxsearch.com/> 官网 download->release
- ii. <http://sphinxsearch.com/docs/current.html> 官网文档

c) 安装

i. 安装依赖包

```
yum install postgresql-libs unixODBC
```

ii. 下载 (源码或者 rpm 包)

查看系统版本 `getconf LONG_BIT`

下载 `wget http://sphinxsearch.com/files/sphinx-2.2.10-release.tar.gz`

iii. 安装操作

```
tar zxvf sphinx-2.2.10-release.tar.gz
```

```
cd sphinx-2.2.10-release
```

```
./configure --prefix=/Web/apps/sphinx //安装目录 ( ./configure --help )
```

```
make
```

```
make install
```

```
cp -a sphinx.conf.dist sphinx.conf
```

//配置文件在 `sphinx/etc/sphinx.conf` 默认不存在 ,可复制 `sphinx.conf.dist`

iv. 生成索引文件

```
sphinx/indexer --config /Web/apps/sphinx/etc/sphinx.conf --all --rotate
```

v. 加入环境变量

```
ln -s /Web/apps/sphinx/bin/indexer /usr/bin/indexer
```

```
ln -s /Web/apps/sphinx/bin/searchd /usr/bin/searchd
```

vi. 启动 sphinx

```
searchd --config /Web/apps/sphinx/etc/sphinx.conf
```

四、sphinx 如何工作

a) sphinx 配置

- i. 配置文件位置 /Web/apps/sphinx/etc/sphinx.conf
- ii. 官方配置文件说明 (12.2. Index configuration options)

<http://sphinxsearch.com/docs/current.html#conf-reference>

b) 利用 indexer 工具生成索引数据

- i. /usr/bin/indexer --config sphinx/etc/sphinx.conf --all
- ii. 报错 indexer: error while loading shared libraries: libmysqlclient.so.18:
cannot open shared object file: No such file or directory

iii. 解决

- 1. 先找到文件 find / -name libmysqlclient.so.18
- 2. ln -s /Web/apps/mysql/lib/libmysqlclient.so.18 /usr/lib/libmysqlclient.so.18
- 3. libmysqlclient.so.18, 软链接放在/usr/lib 下, 若是 64 位放在/usr/lib64 下

iv. 查看索引文件是否生成

- 1. 查看 path = /Web/apps/sphinx/var/data/cdr

c) 索引文件存储的数据种类

- i. .spa 存储文档属性
- ii. .spd 存储每个词 ID 可匹配的文档 ID 列表
- iii. .sph 存储索引头信息
- iv. .spi 存储词列表
- v. .spm 存储 MVA 数据
- vi. .spp 存储每个词的命中列表

d) 使用 searchd 工具

- i. searchd 后台进程, 使用 indexer 工具生成的数据做查询

ii. 执行流程图



iii. 查看进程 `ps -aux | grep searchd`

iv. 查询帮助信息 `searchd --help`

v. 启动 sphinx `searchd --config /Web/apps/sphinx/etc/sphinx.conf`

vi. 关闭 sphinx `killall searchd`

五、查看 sphinx 数据

a) 进入 MySQL 终端查询数据

i. `mysql -h0 -P9306` 指定端口

ii. sphinx 配置文件端口

1. `searchd { listen = 9306:mysql41 }` //与 MySQL 通信端口

b) 查询数据

i. `SELECT * FROM `index_file_name``

ii. 查看配置文件 `index cdr_file_main{ }`

c) 查询指定字段必须在配置文件通过

i. `sql_attr_uint | sql_attr_bool | sql_attr_timestamp` 等

六、理论篇总结

a) 必须掌握配置文件的配置选项

b) 查阅官方手册