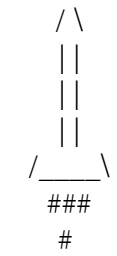
ROCKET SEARCH



Die Hauptverwendungszweg der Software ist das Indizieren von Dateien auf Netzlaufwerken und dessen zur Verfügung stellen dieser Daten via Weboberfläche für den Endanwender.

Ein Anwendungsbeispiel:

an einem Server sind 3 Netzlaufwerke angebunden.

2x Samba, einmal normales Share, und einmal ein Windows DFS share
1x NFS

Filesystem	Size	Used	Avail	Use%	Mounted on
//172.17.190.34/dfs	7.0T	5.3T	1.8T	76%	/root/smb_mount/dfs
172.17.190.6:/usr/sap	466G	244G	222G	53%	/root/smb_mount/sap
//172.17.190.1/docs-team	150G	138G	12G	92%	/root/smb_mount/docs-team

Auf den 3 Netzlaufwerken liegen 15 mio. Dateien in 1.1 mio Verzeichnissen.

```
find "/root/smb_mount/" -noleaf -type f -print  | wc -l
15 182 764
find "/root/smb_mount/" -noleaf -type d -print  | wc -l
1 110 377
```

Alle diese Dateien haben die unterschiedlichsten Ausprägungen aber besonders in den Docs Ordnern, befinden sich überwiegend Microsoft Office Dokumente (Excel & Word) .

Die Redis Datenbank und der auszuführende Server der Indizierung sollten auf einem Physikalischen System liegen.

Die Redis Datenbank starten und in einem Zug konfigurieren.
Es empfiehlt sich dies in einem Startscript zu implementieren.

```
/usr/local/bin/redis-server &
sleep 10
/usr/local/bin/redis-cli config set save ""
/usr/local/bin/redis-cli config set appendonly no
/usr/local/bin/redis-cli config set protected-mode no
```

um Sicherzustellen das keine Datensätze in den SWAP geschrieben werden, diesen abschalten oder auf ein Minimum begrenzen.
Maximale SWAP größe der Partion 2GB .

```
swapoff -a
free -m
             total       used       free     shared    buffers     cached
Mem:        516916      36312     480604         60        888      28435
-/+ buffers/cache:       6988     509928
Swap:            0          0          0
```

oder bei max. 2GB

```
free -m
             total       used       free     shared    buffers     cached
Mem:        516916      36312     480604         60        888      28435
-/+ buffers/cache:       6988     509928
Swap:         2047          0       2047
```

Einige Kernel Parameter für die Redis Datenbank einstellen

```
echo 1 > /proc/sys/vm/overcommit_memory
echo never > /sys/kernel/mm/transparent_hugepage/enabled
echo 1 > /proc/sys/net/ipv4/tcp_tw_reuse
```

Ermitteln wieviele Kerne dem System zur Verfügung stehen.

```
cat /proc/cpuinfo | grep -i processor | wc -l
72
```

Zur Ausführung der Indizierung empfiehlt es sich 75% der zur Verfügung stehenden CPU Kerne zu verwenden.

```
echo "(`cat /proc/cpuinfo | grep -i processor | wc -l` * 75)/100" | bc
```

Die Ausführung des Indizierungsdiestens außerhalb der Kernarbeitszeiten via Cron Job einplanen.

```
#Start 23:59 Uhr jeden Tag#
59 23 * * *     root    wrapper /root/smb_moun/ 54
```

sowie die Bandbreitenbegrenzung auf 5 MBit / Sek. begrenzen.

```
#QoS fuer SMB und NFS
tc qdisc del dev eth0 root
tc qdisc add dev eth0 root handle 1:0 htb default 10
#OUTPUT
tc class add dev eth0 parent 1:0 classid 1:123 htb rate 5mbit ceil 5mbit
tc filter add dev eth0 parent 1: prio 0 protocol ip handle 123 fw flowid 1:123
#INPUT
tc class add dev eth0 parent 1:1 classid 1:456 htb rate 5mbit ceil 5mbit
tc filter add dev eth0 parent 1: prio 0 protocol ip handle 456 fw flowid 1:456
#
tc qdisc show dev eth0
tc class show dev eth0
tc filter show dev eth0
tc -s -d qdisc show dev eth0
#
iptables -F -t mangle
iptables -t mangle -A INPUT -d 172.17.190.78 -s 172.17.190.6 -j MARK --set-mark 456
iptables -t mangle -A INPUT -d 172.17.190.78 -s 172.17.190.34 -j MARK --set-mark 456
iptables -t mangle -A INPUT -d 172.17.190.78 -s 172.17.190.1 -j MARK --set-mark 456
#
#Traffic Controll für SMB
iptables -t mangle -A OUTPUT -p tcp --dport 445 -j MARK --set-mark 123
#Traffic Controll für NFSv4
iptables -t mangle -A OUTPUT -p tcp --dport 2049 -j MARK --set-mark 123
iptables -L -v -n -t mangle
```

Je nach Anzahl der Daten kann der Indizierungsprozess mehrere Stunden oder Tage dauern, abhängig davon was alles indiziert wird. Nach der Indizierung befinden sich nun die Daten komplett im RAM der Redis Datenbank .

Um zu prüfen wie viel Datensätze sich in der Datenbank befinden, folgendes ausführen.

```
redis-cli dbsize
(integer) 24 603

redis-cli info memory
# Memory
used_memory:125 252 320
used_memory_human:119.45M
used_memory_rss:132 771 840
used_memory_rss_human:126.62M
used_memory_peak:125 252 320
used_memory_peak_human:119.45M
used_memory_peak_perc:100.00%
used_memory_overhead:12 790 078
used_memory_startup:786 456
used_memory_dataset:112 462 242
used_memory_dataset_perc:90.36%
total_system_memory:4 008 566 784
total_system_memory_human:3.73G
used_memory_lua:37 888
used_memory_lua_human:37.00K
maxmemory:0
maxmemory_human:0B
maxmemory_policy:noeviction
mem_fragmentation_ratio:1.06
mem_allocator:jemalloc-4.0.3
active_defrag_running:0
lazyfree_pending_objects:0
```

Da während der Indizierung kein Backup der Daten aus der Datenbank erfolgt, sollte man nach Beendigung der Indizierung ein Backup manuell durchführen.

```
redis-cli save
```

Der dump.rdb Datenbank Backup File befindet sich dort, wo die ausführbare Datei „redis-server" gestartet wurde. Da diese über ein Startscript gestartet wurden, liegt es in diesem Fall unter /root/scripte

```
~/scripte # ls -altr
total 95220
-rwxrwxrwx 1 root root      986 Jun 29 14:53 master_start_script.sh
drwx------ 1 root root      558 Jun 29 14:53 ..
-rw-r--r-- 1 root root 97497969 Jun 29 14:55 dump.rdb
drwxr-xr-x 1 root root       60 Jun 29 14:56 .
```

Den Backup File sollte man nun sicher verwahren.

Der Bereich für das Backup zur Indizierung der Daten ist nun abgeschlossen. Hier geht es weiter mit dem Webfrontend.

Dieses sollte aus Performance gründen nicht auf dem Rocket-Search System Server liegen, sonder separat. Als Web-Server sollte ein nginx zum Einsatz kommen.

Nginx mit folgendem Kommando starten:

nginx
sowie php-cgi: „nohup php-cgi -b 127.0.0.1:9000 &“

Im folgenden Schritt mittels Browsers, empfehlenswert Firefox, auf den Nginx Server zugreifen.

http://<ngix webserver>

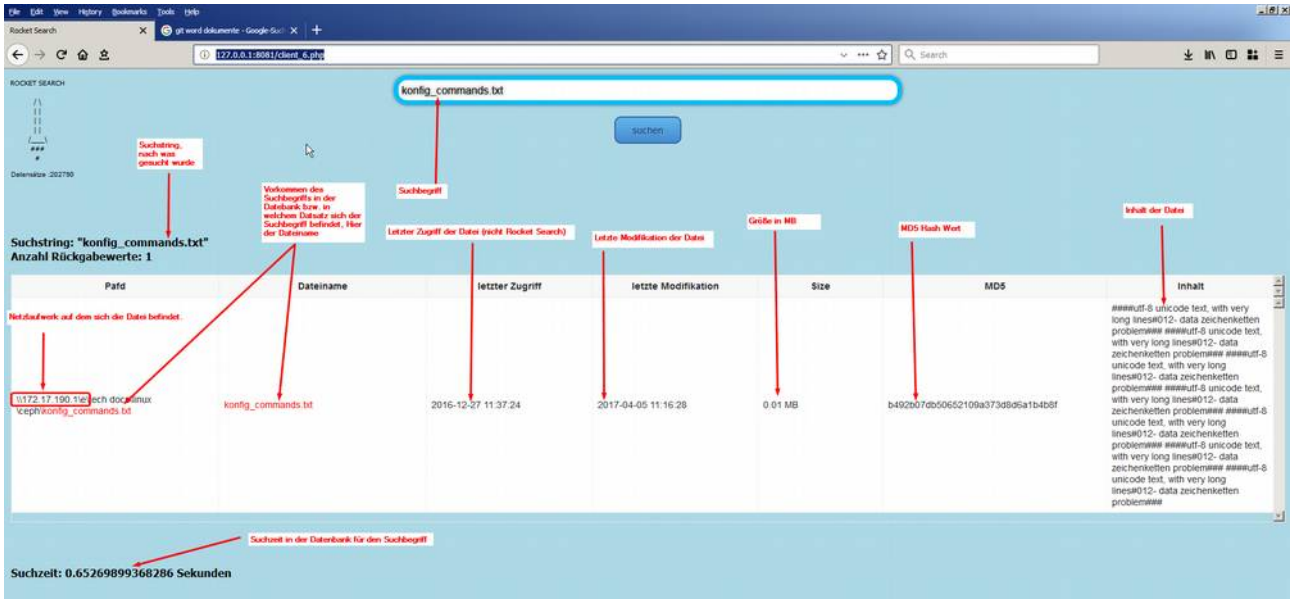
Das WebUI ist in zwei größer Bereich aufgeteilt. Rechts und mittig die Suche sowie links der Statistik Bereich.

Um etwas zu suchen, einfach einen Suchbegriff im Textfeld eingeben und auf suchen klicken.
Möglichen Suchbegriffe wären zum Beispiel:

Der Name einer Datei oder eines Verzeichnisses
Dateiendungen
Erstelldatum einer Datei
Versionsnummer einer EXE oder DLL Datei
Inhalt eines ASCII Files

MD5 Hash Wert einer Datei

Das folgende Schaubild zeigt die einzelnen Ergebnisse der Suche.



Mit z.B. Doppelklick auf den Netzlaufwerkpfad unter „Pfad“ , ist es möglich mittels Kopieren und Einfügen die Datei ganz einfach zu starten oder zu betrachten.