

Fili Wiese / Kaspar Szymanski

# SCREAMING FROG IN DER GOOGLE CLOUD

## DER AUTOR



**Fili Wiese** war bis 2012 sieben Jahre bei verschiedenen Teams bei Google tätig und arbeitete unter anderem an Webspam und Klickspam Projekten. Heute arbeitet heute als unabhängiger SEO Berater mit dem Schwerpunkt technische Optimierung, Link Analyse und Web Development.

## DER AUTOR



**Kaspar Szymanski** war von 2006 bis 2013 für das Google Search Quality Team tätig. In dieser Zeit war er mit dem Aufspüren von Webspam sowie mit der Kommunikation mit Webmastern in EMEA betraut. Aktuell arbeitet er als unabhängiger SEO Berater mit den Schwerpunkten Risikomanagement, Backlink Analyse und Penalty Recovery.

Screaming Frog ist eines der wenigen Tools, auf die ein ernsthaft analytisch arbeitender SEO nicht verzichten kann. Zweifellos ist es ein hervorragendes Tool zum Crawlen von Websites. Bei umfangreicheren Websites stößt man damit allerdings schnell an Grenzen, weil dann der Hauptspeicher des Rechners plötzlich nicht mehr ausreicht. Was tut man, wenn man keine Möglichkeit für Speichererweiterungen mehr hat? Die beiden ehemaligen Google-Mitarbeiter und SEO-Experten Kaspar Szymanski und Fili Wiese zeigen eine Lösung auf: Man lässt Screaming Frog in der Cloud bzw. einfach über die Google-Infrastruktur laufen. Das ist technisch gesehen nicht ganz einfach – aber mit guten Computerkenntnissen und einer derart kompetenten Hilfestellung durchaus zu bewerkstelligen. Die Belohnung für die Installationsmühen sind nahezu unbeschränkte Systemressourcen.

## Screaming Frog SEO Spider

Aus Sicht eines SEO ist „Screaming Frog SEO Spider“ einer der besten Crawler zum Durchsuchen und Analysieren von Websites. Entwickelt von einer Agentur namens Screaming Frog in UK, ist Screaming Frog SEO Spider einer der wenigen – aber nicht der einzige – Crawler, der auch auf einem Debian-basierten System wie Ubuntu arbeitet. Das Tool kann für verschiedenste Zwecke eingesetzt werden wie bspw. [OnPage\\*](#)-Analysen und

Untersuchungen von Backlink-Profilen. Mehr Informationen dazu findet man unter [www.seerinteractive.com/blog/screaming-frog-guide](http://www.seerinteractive.com/blog/screaming-frog-guide).

Die größte Herausforderung beim Einsatz von Screaming Frog SEO Spider ist der enorme RAM-Speicherbedarf, um große Webseiten oder URL-Listen zu crawlen. Zwar arbeitet das Entwicklungsteam von Screaming Frog an verschiedenen Updates, aber man kann auch das enorme RAM-Speicherpotenzial der Google-Infrastruktur

\*siehe Glossar auf Seite 148-150

```

fili@daedalus: ~
File Edit View Search Terminal Help
fili@daedalus:~$ gcloud version
Google Cloud SDK 0.9.17
Copyright Google Inc. 2013.

bq 2.0.18
bq-nix 2.0.18
core 2014.02.26
core-nix 2014.01.27
gcutil 1.13.0
gcutil-nix 1.13.0
gsutil 3.42
gsutil-nix 3.42
sql 2014.02.25

fili@daedalus:~$ █

```

Abb. 1: Verification the installation of Google Cloud SDK

nutzen, um den Screaming Frog SEO Spider laufen zu lassen. Hier kommt die Google Compute Engine ins Spiel.

## Die Google Compute Engine

Neu für manche Anwender mag vielleicht sein, dass Google es tatsächlich erlaubt, große Workloads auf virtuellen Maschinen der Google-Infrastruktur laufen zu lassen. Dieser Vorgang, der die Möglichkeit eröffnet, Computer-Ressourcen anwendungsorientiert anzumieten, wird auch als „Google Compute Engine“ bezeichnet. Aktuell noch im Entwicklungsstadium, ist die Google Compute Engine bereits jetzt ein ernst zu nehmen der Wettbewerber beim Angebot von Cloud-Computing-Ressourcen, verglichen mit Services wie Amazon Cloud und Microsoft Azure Cloud. Aufgrund der attraktiven Preisgestaltung auf Wettbewerbsniveau oder darunter bietet sich die Google Compute Engine als hervorragende Alternative zum bereits etablierten Amazon-Cloud-Service an.

Und wenn wir als Autoren ganz ehrlich sind: Als SEO und ehemalige Google-Mitarbeiter finden wir die Idee großartig, die Google-Infrastruktur für das Crawlen von Webseiten für uns zu nutzen.

## Installieren der Google Cloud SDK

Für den Start muss man zunächst die Google Cloud SDK lokal auf dem Computer installieren. Dieser Prozess ist etwas zeitaufwendig, muss aber nur einmal durchgeführt werden. Da es einfacher ist, diesen Prozess mit einer Linux-Distribution wie Ubuntu laufen zu lassen, werden die folgenden Schritte auf Linux-Basis vorgestellt.

Die Beschreibung einer Installation der Google Cloud SDK auf einem lokalen Windows-Rechner ist zu finden unter <https://developers.google.com/compute/>.

Zuerst wird ein Fenster geöffnet und sichergestellt, dass „curl“ installiert ist. Dies kann getestet werden, indem man „curl“ in das Fenster eintippt und darauf wartet, ob der Vorschlag erscheint, das Programm zu installieren. Falls nötig, ist „curl“ dann zu installieren.

Nachdem „curl“ sauber installiert wurde, sind die folgenden Befehle in das Fenster (aus dem Home-Verzeichnis) einzugeben. Daraufhin erfolgt der Download und die Installation der Google Cloud SDK.

```

$ curl https://dl.google.com/dl/cloudsdk/
release/install_google_cloud_sdk.bash | bash

```

Der nächste Schritt ist, das Terminal wissen zu lassen, dass Google Cloud SDK installiert wurde. Es gibt zwei Möglichkeiten, von denen die einfachste ist, das Fenster zu schließen und dann wieder zu öffnen. Alternativ kann man auch den folgenden Befehl eingeben (welcher vermeidet, das Fenster neu zu starten zu müssen):

```

$ source ~/.<bash-profile-file>

```

Danach kann mit dem folgenden Befehl überprüft werden, dass die Google Cloud SDK tatsächlich installiert ist:

```

$ gcloud version

```

Wenn keine Fehlermeldung erscheint und eine Versionsnummer sowie eine Liste der verschiedenen installierten Tools angezeigt wird, ist die Google Cloud SDK auf Ihrem System installiert und funktioniert.

Der nächste Schritt ist, den Computer für den Google-Cloud-Dienst zu authentifizieren. Damit ist es möglich, Befehle an Google Cloud zu schicken, um verschiedene Google-Cloud-Services zu verwalten. Dazu wird folgender Befehl eingegeben:

```

$ gcloud auth login

```

In Abhängigkeit davon, ob ein Browser auf dem lokalen Computer installiert ist, wird ein neues Browser-Fenster geöffnet. Dort wird man aufgefordert, Google Cloud SDK die Zugangserlaubnis für das eigene Google-Konto zu geben (möglicherweise wird man zuerst aufgefordert, sich in sein Google-Konto einzuloggen.)

Alternativ kann auch die Aufforderung erfolgen, einen Link (vom Fenster) zu kopieren, um den Vorgang abzuschließen. Bei erfolgreicher Registrierung erscheint folgende Bestätigung: „Sie sind jetzt mit dem Google Cloud SDK authentifiziert.“

Währenddessen wird man aufgefordert, die Google-Cloud-Projekt-ID einzugeben. Hierzu drückt man einfach ENTER. Wenn alles gut gegangen ist, dann erhält man eine Nachricht, dass man jetzt in seinem Google-Konto eingeloggt ist.

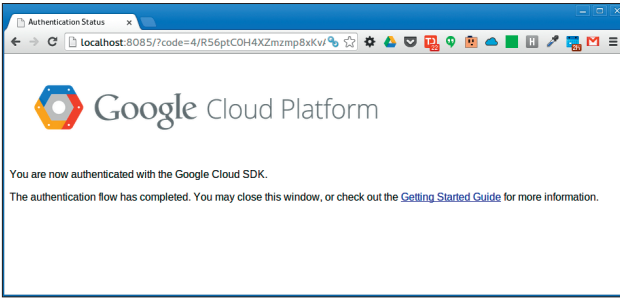


Abb. 2: Die Bestätigung der Google Cloud SDK in der Google Cloud Platform

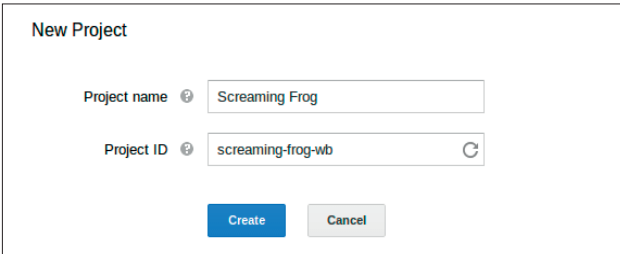


Abb. 3: Ein neues Projekt in der Google Cloud Console anlegen

### Google Developer Console

Der nächste Schritt besteht darin, unter <https://cloud.google.com/console> zur Google Cloud Console for Developers zu gehen. Dies ist die Website für die Verwaltung Ihrer Google-Cloud-Dienste ähnlich wie die Google Compute Engine.

Zu Beginn muss ein neues Projekt angelegt werden. Hierfür klickt man auf den roten Knopf im oberen Bereich der Seite „Create Project“. Ein Pop-up-Fenster erscheint, in dem der Projektname eingegeben und die Projekt-ID definiert wird. Der Projektname ist nicht so wichtig, da er nur in der Google Cloud Console verwendet wird. Am einfachsten gibt man „Screaming Frog“ ein.

Die Projekt-ID ist sehr wichtig, da sie das Projekt unter allen Google-Cloud-Services-Benutzern eindeutig identifiziert. Wenn man nicht mit dem Standardvorschlag von Google arbeitet (Klick auf den kleinen Pfeil auf der rechten Seite im Eingabefeld für weitere Vorschläge von Google), kann es einige Anstrengung kosten, um eine verfügbare und eindeutige Projekt-ID zu finden. Für das hier vorgestellte Projekt wird „screaming-frog-wb“ verwendet.

Sobald man auf den Button „Erstellen“ klickt, wird man zur Übersichtsseite des Projekts umgeleitet, in diesem Fall <https://console.developers.google.com/project/apps-screaming-frog-wb> (zu beachten ist, dass die Projekt-ID hier in der URL verwendet wird).

Jetzt kommt ein wichtiger Schritt: Die Abrechnungs- und Zahlungsmöglichkeit muss eingerichtet werden. Google Compute Engine hat keine freien Kontingente; für die Nutzung muss die Möglichkeit einer Bezahlung eingeräumt werden. Informationen zur Preisgestaltung von Google Compute Engine finden sich unter <https://developers.google.com/compute/pricing>.

Um die Bezahlung zu aktivieren, klickt man auf „Settings“ in der linken Seitenleiste der Projekt-Übersichtsseite. Die erste

Option, die sichtbar wird, erscheint als grauer Button „Enable billing“. Hier klickt man und sollte seine Kreditkarte bereithalten. Dann wählt man das richtige Land und gibt Adresse, Steuereinformationen (falls vorhanden), Telefonnummer und Name sowie die Kreditkartendaten ein. Sobald dieser Schritt abgeschlossen ist, kann die Google Compute Engine gestartet werden.

**Hinweis:** Wenn einmal „Enable billing“ aktiviert wurde und die Google Cloud Services genutzt wurden, erscheint auf der Projektübersichtsseite ein Link in der rechten oberen Ecke, der Hinweise auf die geschätzten Kosten für den laufenden Monat gibt.

Alle Schritte bis jetzt bezogen sich nur auf das Einrichten eines Projekts und der Software auf dem steuernden Computer. Die meisten dieser Schritte brauchen nicht wiederholt werden, es sei denn, man wechselt den Computer oder möchte neue Projekte einrichten.

### Instanz der eigenen Google Compute Engine starten

Das Fenster wird erneut geöffnet und die folgenden Befehle eingegeben, wobei <project-id> durch die aktuelle Project-ID zu ersetzen ist; für das vorliegende Beispiel wäre das screaming-frog-wb:

```
$ gcloud config set project <project id>
```

Alles, was nun mit dem Google Cloud SDK ausgeführt wird, ist jetzt Teil des Projektes Screaming Frog, welches auch die initiierte Bezahlung beinhaltet.

Zur Bestätigung, dass noch nichts gestartet wurde, führt man den Befehl aus:

```
$ gcutil listinstances
```

Dieser Befehl sollte eine leere Tabelle anzeigen.

Nun kann man dazu übergehen, eine neue Instanz zu generieren, indem man folgenden Befehl ausführen lässt:

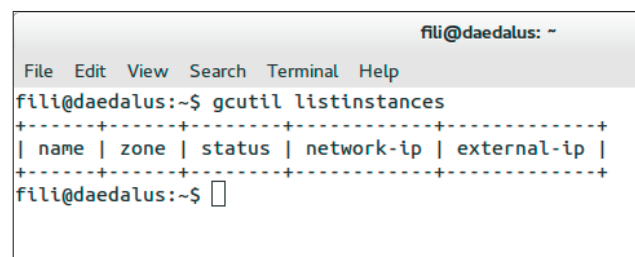


Abb. 4: Bestätigung, dass keine Compute-Engine-Instanz läuft

```
$ gcutil addinstance screaming-frog-test
```

Der aufmerksame Leser wird bemerkt haben, dass der „-wb“-Teil als Befehl ausgelassen wurde. Das ist nötig, weil „screaming-frog-test“ eine weitere eindeutige Identifizierung für das Beispiel im Rahmen des Projektes „screaming-frog-wb“ ist.

Im Fenster wird man nun aufgefordert, einen Bereich, einen Gerätetyp und ein Bild auszuwählen. Hier kommt man mit der Option us-central1-a zone, f1-micro machine type (am günstigsten) und debian-7-wheezy image weiter. Die Instanz ist nun eingerichtet. Nach Fertigstellung kann folgender Befehl ausgeführt werden, um nach dem Login SSH zu nutzen (Kommandozeile):

```
$ gcutil ssh screaming-frog-test
```

**Hinweis: Man könnte aufgefordert werden, den SSH-Schlüssel einzurichten. In diesem Fall folgt man einfach den Anweisungen und verwendet eine Passphrase, an die man sich gut erinnern kann.**

Herzlichen Glückwunsch! Jetzt ist man mit der virtuellen Google-Infrastruktur verbunden. Die Bestätigung erfolgt, indem man zu seinem Projekt in der Google Cloud Console geht oder den folgenden Befehl ausführt:

```
$ gcutil listinstances
```

In dieser Phase beendet man die gestartete Instanz wieder, und zwar mit dem folgenden Befehl:

```
$ exit
```

Jetzt wird man abgemeldet und löst die Verbindung zwischen dem Computer und der Instanz bei Google. Dann springt man zur Google Cloud Console, wählt sein Projekt aus, die Compute Engine, die VM-Instanzen, klickt auf den Link screaming-frog-test und geht zum unteren Rand der Seite. Hier kann man auf den „Delete“-Button klicken, um die Instanz zu löschen. Wenn man darauf klickt, darf man nicht vergessen, auch die Boot-Disk screaming-frog-test zu löschen.

Alternativ – um die Instanz wieder herunterzufahren – gibt man den folgenden Befehl ein:

```
$ gcutil deleteinstance screaming-frog-test
--delete_boot_pd
```

Man wird nun aufgefordert zu bestätigen, dass die Instanz

gelöscht werden soll. Nach dieser Bestätigung wird die Google Compute Engine versuchen, die virtuelle Instanz und die Boot-Disk zu löschen. Man kann erneut bestätigen, dass die Instanz heruntergefahren wird (dafür fallen auch keinerlei Kosten an), indem man folgenden Befehl eingibt:

```
$ gcutil listinstances
```

**Hinweis: Manchmal kann es bei den Google-Cloud-Services zu einer Verzögerung und damit zu einem Timeout kommen. Wenn das passiert, kann der Löschvorgang in der Google Cloud Console nachvollzogen werden.**

## Einrichtung der Screaming-Frog-Instanz

Nun, da das Google-Cloud-Projekt eingerichtet ist und die grundlegenden Befehle, um mit Google-Compute-Engine-Instanzen arbeiten zu können, bekannt sind, ist es Zeit, eine Instanz mit dem Screaming Frog Spider einzurichten.

Zuerst wird eine neue Instanz durch Ausführen der folgenden Befehle erzeugt (nachfolgend wird „screaming frog“ als eindeutige Kennung für die Instanz verwendet):

```
$ gcutil addinstance screaming-frog --service_
account_scopes=storage-rw
```

Es steht einem frei, jede regionale Zone auszuwählen; man sollte sich allerdings bewusst sein, dass die europäischen Zonen etwas teurer sein können. Man wählt ein Gerät mit ausreichend RAM (die Autoren neigen dabei zu „n1-Standard-8“) und debian-7-wheezy Bild. Letzteres ist sehr wichtig!

Der aufmerksame Leser mag auch das zusätzliche Flag für service\_account\_scopes im vorherigen Befehl bemerkt haben. Dieses Flag ermöglicht es, Installationen später zu speichern, und spart jedes Mal Zeit, wenn Screaming Frog SEO Spider auf der Google Compute Engine zukünftig zum Einsatz kommen soll.

```
$ gcutil ssh screaming-frog
```

Hat man sich in der Instanz angemeldet, ist es nun nötig, zur eigenen Root zu wechseln, indem man folgenden Befehl verwendet:

```
$ sudo -s
```

Jetzt ist ein Update der Software-Packages notwendig. Man führt den folgenden Befehl aus:

```
$ apt-get update
```

Dann gibt man den nächsten Befehl ein, um die erforderlichen Programme zu installieren:

```
$ apt-get install tightvncserver xfce4 xfce4-goodies xdg-utils openjdk-6-jre
```

Das dauert einige Minuten und installiert einen VNC-Server sowie ein minimalistisches Graphical User Interface, das nur geringe Ressourcen verbraucht. Wenn man nach einer Keyboard-Konfiguration gefragt wird, wählt man einfach die Voreinstellung aus (man verwendet Tab auf dem Keyboard, um zur „OK“-Option zu gelangen und „Enter“ auszuführen).

```
$ dpkg-reconfigure locales
```

Nun wird man aufgefordert, Locale auszuwählen. Die einfachste (aber auch zeitaufwendigste) Methode ist, „All Locales“ auszuwählen (man verwendet Tab, um zur „OK“-Option zu kommen). Dann wählt man die Voreinstellung „None“ als Standard für die Systemumgebung. Dieser Prozess wird abermals einige Minuten in Anspruch nehmen.

Sobald der Prozess einmal abgeschlossen ist, kann man dazu übergehen, einen anderen User zum System hinzuzufügen mit dem Namen „vnc“:

```
$ adduser vnc
```

Nach der entsprechenden Aufforderung wählt man ein sicheres Passwort. Man kann alles andere überspringen, indem man die Standardeinstellung mit „Enter“ bestätigt. Mit „Y“ bestätigt man die Richtigkeit der Informationen.

Dann muss man ein neues Passwort für den User einrichten, das übrigens maximal acht Zeichen lang sein darf. Als Erstes wechselt man zu einem neuen User durch folgenden Befehl:

```
$ su vnc
```

Es folgt das Ausführen des nächsten Befehls:

```
$ vncpasswd
```

Wenn man gefragt wird, ob man ein view-only password eingeben möchte, gibt man (n) für Nein ein. Dieser Vorgang erzeugt ein neues Verzeichnis im Home-Directory des VCN-Nutzers und erzeugt ein neues Passwort, das später für eine VNC-Verbindung zur Instanz verwendet wird.

## Einrichten des Start-up-Scripts

Nachdem der VCN-Nutzer eingerichtet wurde, müssen noch einige Start-up-Scripts installiert werden, die dafür sorgen, dass der VNC-Server jedes Mal die Instanz startet bzw. dass ein Neustart durchgeführt wird. Zunächst wechselt man dazu zurück zum Root User durch das Eingeben der folgenden Befehle:

```
$ exit
```

Jetzt wird das erste Start-up-Script heruntergeladen, indem folgender Befehl ausgeführt wird:

```
$ wget http://fliwiese.com/files/vncserver -O /etc/init.d/vncserver
```

Das zweite Start-up-Script folgt über den nächsten Befehl.

```
$ wget http://fliwiese.com/files/xstartup -O /home/vnc/.vnc/xstartup
```

Nachdem die Start-up-Scripts heruntergeladen und installiert wurden, lässt man den VCN-Server seinen Job ausführen mit den folgenden Befehlen:

```
$ chown -R vnc. /home/vnc/.vnc && chmod +x /home/vnc/.vnc/xstartup
$ sed -i ,s/allowed_users.*/allowed_users=anybody/g' /etc/X11/Xwrapper.config
$ chmod +x /etc/init.d/vncserver
```

Die Instanz wird neu gestartet mit folgendem Befehl:

```
$ reboot
```

Die SSH-Verbindung wird zu dieser Zeit beendet, was eine oder zwei Minuten dauern kann. Einen erneuten Zugang zur Instanz durch SSH erhält man durch folgenden Befehl:

```
$ gcutil ssh screaming-frog
```

Man wechselt zurück zum Root User beim Ausführen des nachfolgenden Befehls:

```
$ sudo -s
```

Jetzt kann der VNC-Service über folgende Befehle gestartet werden:

```
$ update-rc.d vncserver defaults
$ service vncserver start
```



Herzlichen Glückwunsch! Jetzt kann jedes VNC-fähige Programm für einen Zugang zur Instanz über die VNC-Verbindung benutzt werden.

## Installation von Screaming Frog SEO Spider und Dropbox

Vor einer Verbindung durch VNC beendet man den Installationsprozess als Erstes mit dem Installieren von Frog SEO Spider und Dropbox. Zum Download von Screaming Frog Spider gibt man folgenden Befehl ein:

```
$ wget http://www.screamingfrog.co.uk/
products/seo-spider/screamingfrogseospi-
der_2.22_all.deb
```

Für die Installation von Screaming Frog SEO Spider für alle Nutzer verwendet man diesen Befehl:

```
$ dpkg -i screamingfrogseospider_2.22_all.deb
```

**Hinweis:** Eine neuere Version von Screaming Frog für Ubuntu könnte zeitgleich mit diesem Artikel erscheinen. Sollte dies der Fall sein, ist die neue URL zur letzten Ubuntu-Version auf <http://www.screamingfrog.co.uk/seo-spider/> zu finden.

Weiter zum Installationsprozess von Dropbox. Der Grund, weshalb man Dropbox installiert, ist die Möglichkeit eines File-Transfers von der Instanz zum Computer. Dropbox ist ein hervorragendes Tool für diesen Zweck. Hierfür wird zunächst das Dropbox-Installation-File für Debian heruntergeladen mit folgendem Befehl:

```
$ wget https://www.dropbox.com/download?d-
l=packages/debian/dropbox_1.6.0_amd64.deb -O
dropbox.deb
```

Man installiert Dropbox für den VNC-Nutzer durch das Ausführen des folgenden Befehls:

```
$ dpkg -i dropbox.deb
```

Dies wird aufgrund einer Abhängigkeit – genannt „python-gtk2“ – einen Fehler erzeugen. Durch die Eingabe des folgenden Befehls kann diese Fehlermeldung umgangen werden:

```
$ apt-get -f install
```

Dann ist die Dropbox-Installation erfolgreich abgeschlossen. Weiter geht es mit der Verbindung zum VNC-Server.

## Verbindung mit VNC

Um herauszufinden, mit welcher IP-Adresse die Instanz zuzugreifen versucht, loggt man sich aus der SSH-Verbindung aus und führt den folgenden Befehl aus, sodass die externe IP-Adresse in der Tabelle gelistet wird.

```
$ gutil listinstances
```

Bevor man den nächsten Schritt geht, müssen die Firewall-Vorgaben der Instanz auf den aktuellen Stand gebracht werden. Dafür geht man in die Google Cloud Console im Browser, wählt sein Projekt aus, die Compute Engine, das Netzwerk in der linken Seitenleiste und klickt auf den „Default“-Link. Hier muss man eine neue Vorgabe hinzufügen. Man findet die Firewalls auf der Seite, klickt auf den Link „Create new“ und verwendet nachfolgende Details zum Ausfüllen der Felder:

```
Name: vnc
Protocols & ports: tcp:5800,5900-5909
```

Man verwendet die Standardeinstellungen für die verbleibenden Felder und klickt den blauen „Create“-Button. Eine neue Firewall-Vorgabe wird erzeugt, die den Zugang zum VNC-Server ermöglicht.

Zur Verbindung mit der Instanz über VNC auf Ubuntu kann man das Programm Remmina ausprobieren, indem man folgende Details eingibt:

```
hostname: <external-ip>:5901
password: <the-8-character-password>
```

Sobald die VNC-Verbindung eingerichtet ist, erscheint ein Pop-up-Fenster auf dem Desktop, indem die Option „Use default config“ ausgewählt wird.

Dann geht man zum Anwendungsmenü in der oberen rechten Ecke und wählt „Internet“ und „Dropbox“ aus. Daraufhin wird Dropbox für den VNC-User heruntergeladen mit der Aufforderung, entweder einen neuen Dropbox-Account anzulegen oder sich in einen bestehenden Account einzuloggen. Man erstellt einen neuen Account, um zu vermeiden, dass bereits bestehende Dropbox-Files in die neue Instanz heruntergeladen werden. Sobald das Dropbox-Set-up und die Registrierung abgeschlossen sind, eröffnet man einen neuen Folder innerhalb der Dropbox und teilt diesen mit dem bestehenden Account. Nun können Files zwischen der Instanz und dem Computer via Dropbox ausgetauscht werden.

Dann kann man eine Verknüpfung auf dem Desktop zum Screaming Frog SEO Spider erstellen: Rechtsklick auf den

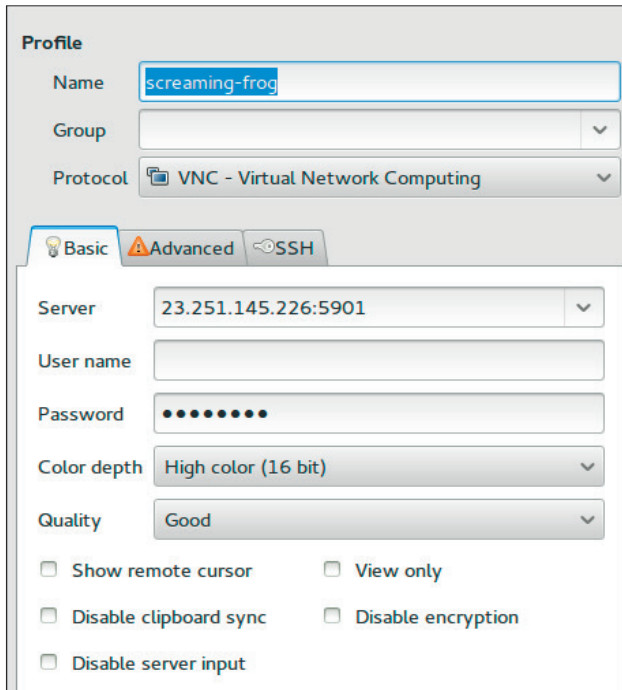


Abb. 5: Beispielkonfiguration des VNC-Programms

Desktop-Hintergrund, „Create Launcher“ auswählen und folgende Details in die Felder eingeben:

```
Name: Screaming Frog
Command: screamingfrogseospider %f
```

Anschließend klickt man auf den „Create“-Button. Eine neue Verknüpfung ist entstanden und wird auf dem Desktop angezeigt. Ein Doppelklick startet den Screaming Frog SEO Spider. An diesem Punkt ist es hilfreich, die Lizenzinformationen noch vor dem Schließen des Programmes erneut einzugeben.

**Hinweis:** Bei der Eingabe dieses Namens erhält man möglicherweise einen Vorschlag, um „Create Launcher Screaming Frog SE ...“ auszuwählen. In diesem Fall wählt man diese Option aus, auch durch Klick auf den ICON-Button, bevor man den „Create“-Button klickt. Nach Auswahl der Vorschläge kann für das Startprogramm das Standard-ICON Screaming Frog SEO Spider gewählt werden

Der Installationsprozess ist damit fast abgeschlossen. Als weiterer Schritt erfolgt die Anpassung des zugewiesenen Speichers zum RAM der Instanz, den wir zur Verfügung haben. Dieser Prozess funktioniert nur, wenn Screaming Frog SEO mindestens einmal durch VNC gestartet wurde. Man geht in das Fenster zurück und verbindet sich über SSH erneut mit dem folgenden Befehl:

```
$ gcutil ssh screaming-frog
```

Dann wechselt man den User mit folgendem Befehl:

```
$ sudo -s
$ su vnc
```

Jetzt öffnet man das Screaming-Frog-SEO-Spider-Konfigurations-File für den zugewiesenen Speicher mit dem Befehl:

```
$ pico /home/vnc/.screamingfrogseospider
```

Abhängig vom Gerätetyp, der für die Instanz ausgewählt wurde, ändert man die Zahl 512 in eine Zahl, die näher an dem maximalen RAM der Instanz liegt. Wenn bspw. n1-standard-8 benutzt wird, dann ist der verfügbare RAM 30 GB. In diesem Fall ersetzt man die Nummer 512 durch 29000. Dann schließt und speichert man das File durch Drücken von Ctrl-X und bestätigt die Frage, ob man den geänderten Pufferspeicher speichern möchte, mit Y. Zukünftig, wann immer der Gerätetyp der Instanz kleiner ist, muss sichergestellt werden, dass die Nummer auf eine Größe unterhalb des verfügbaren RAM angepasst wird.

Jetzt sind sowohl VNC als auch Screaming Frog SEO Spider und Dropbox eingerichtet.

Nun muss noch sichergestellt werden, dass all diese Schritte nicht jedes Mal wiederholt werden müssen, wenn man eine andere Instanz startet. Dazu speichert man den bisherigen Prozess ab.

## Speichern der Instanz

Zur Speicherung bzw. Sicherung des Ganzen loggt man sich zunächst über den Computer bei der Instanz ein, indem man den folgenden Befehl verwendet:

```
$ gcutil ssh screaming-frog
```

Dann wird der nachfolgende Befehl ausgeführt, um den Back-up-Prozess zu starten:

```
$ sudo gcimagebundle -d /dev/sda -o /tmp/
--log_file=/tmp/abc.log
```

Dieser Befehl erzeugt ein Image von allen Einstellungen und Programmen, die in den vorangegangenen Schritten installiert und eingegeben wurden. Die Ausgabe dieses Befehls zeigt eine lange hexadezimale Nummer, die Name und Speicherort des neu erzeugten Images wiedergibt, wie z. B.:

```
/tmp/<long-hex-number>.image.tar.gz
```

Durch Copy & Paste sichert man sich diese lange Hex-Nummer vorübergehend an irgendeinem Ort, da diese für die nächs-



ten Schritte benötigt wird.

Jetzt, da ein Back-up-Image erstellt wurde, muss dieses auch im Google Cloud Storage gespeichert werden. Nach Authentifizierung und Konfiguration öffnet man den Zugang zum Google Cloud Storage folgendermaßen:

```
$ gsutil config
```

Man folgt den Anweisungen auf dem Bildschirm und öffnet ein neues Browserfenster mit der zur Verfügung gestellten URL, akzeptiert die Zugangsbedingungen und kopiert den Authentifizierungscode zurück in das Eingabefeld der URL. Dann bestätigt man die Projekt-ID – in diesem Fall screaming-frog-wb – und drückt ENTER.

Jetzt erzeugt man einen neuen Bucket in der Google Cloud Storage mit einem eindeutigen Namen durch folgenden Befehl:

```
$ gsutil mb gs://<bucket-name>
```

**Hinweis: Man ersetzt <bucket-name> durch einen eindeutigen Namen, der einzigartig innerhalb aller Google Storage Buckets ist; es erfordert einige Kreativität, hier einen verfügbaren Namen zu finden.**

Der nächste Schritt besteht darin, das Image in den Google Cloud Storage Bucket zu kopieren mit dem Befehl:

```
$ gsutil cp /tmp/<long-hex-number>.image.tar.gz
gs://<bucket-name>
```

**Hinweis: Das Update der <log-hex-number> und des <bucket-name> muss über den vorangegangenen Befehl erfolgt sein!**

Sobald der Kopiervorgang des Back-up-Images in den Google Cloud Storage Bucket abgeschlossen ist, loggt man sich bei der SSH-Verbindung mit der Instanz durch den folgenden Befehl wieder aus:

```
$ exit
```

Um das benutzerdefinierte Back-up-Image zur Sammlung des Google-Cloud-Projektes hinzuzufügen, gibt man ein:

```
$ gcutil --project=screaming-frog-wb addimage
screaming-frog-image gs://
<bucket-name>/<long-hex-number>.image.tar.gz
```

Sobald dieser Prozess abgeschlossen ist, ist das Back-up-Image sicher im Google Cloud Storage gespeichert und inner-

halb des Projektes beim nächsten Mal, wann immer eine Instanz hergestellt wird, verfügbar. Falls notwendig, kann das Image wieder gelöscht werden, durch die Google Cloud Console oder durch folgenden Befehl:

```
$ gcutil --project=screaming-frog-wb deletei-
mage screaming-frog-image
```

Wahlweise kann man überprüfen, dass das Back-up-Image erfolgreich erzeugt wurde, indem man zur Google Cloud Console geht, das Projekt, die Compute Engine und Images in der linken Seitenleiste auswählt, wobei das Image „screaming-frog-image“ in der Liste der verfügbaren Images erscheinen sollte.

An diesem Punkt ist alles konfiguriert und gesichert, sodass das aktuelle Beispiel mit dem Befehl deaktiviert werden kann:

```
$ gcutil deleteinstance screaming-frog
--delete_boot_pd
```

Dies wird die Instanz deaktivieren und die Festplatte löschen, was zusätzliche Kosten vermeidet, bis wir erneut Screaming Frog SEO Spider oder Google Compute Engine nutzen. Die Bestätigung hierfür erfolgt über den Befehl:

```
$ gcutil listinstances
```

Als Ergebnis sieht man anschließend wieder die leere Tabelle.

## Die fertig konfigurierte Instanz ist jetzt nutzbar

Wenn man den Screaming Frog SEO Spider auf der Google Compute Engine später operativ nutzen möchte, öffnet man das Fenster auf seinem Computer und gibt ein:

```
$ gcutil addinstance screaming-frog
```

Wenn man aufgefordert wird, eine Zone auszuwählen (obwohl man sich bewusst sein muss, dass die europäischen Zonen etwas teurer sind), wählt man eine Maschine mit ausreichend RAM (vorzugsweise n1-standard8) und wählt „screaming-frog-image“. Sobald die Instanz geöffnet ist und läuft, merkt bzw. notiert man sich die der Instanz zugewiesene externe IP-Adresse. Dann startet man das VNC-Programm wie bspw. Remmina und schafft eine Verbindung zur Instanz, indem man die externe IP-Adresse auf Port 5901 verwendet und das Passwort eingibt. Nun steht dem Start des Screaming Frog nichts mehr entgegen und die Probleme mit zu wenig Speicherplatz gehören ab jetzt der Vergangenheit an.¶