# 使用Go和AWS Lambda进行简单的网站文本抓取



最近我需要知道某些网站何时更新了特定的文本。我决定使用AWS Lambda来节省托管服务器的成本，并使用Go是因为它速度快，而且也是因为它是awslambda上支持的语言之一。我还使用AWS SES在发现结果时向我发送电子邮件通知。

下面我将向您展示如何编译Go脚本，设置AWS Lambda函数，并配置cron类型的作业以每小时运行一次脚本。

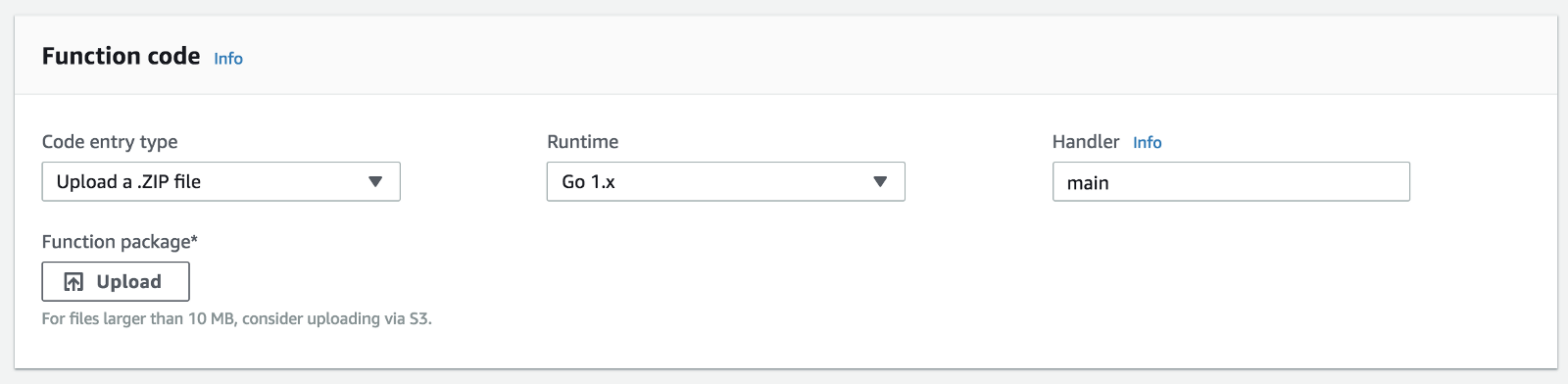
首先，克隆包含脚本的repo。

<https://github.com/aaronvb/aws_lambda_go_scraper>

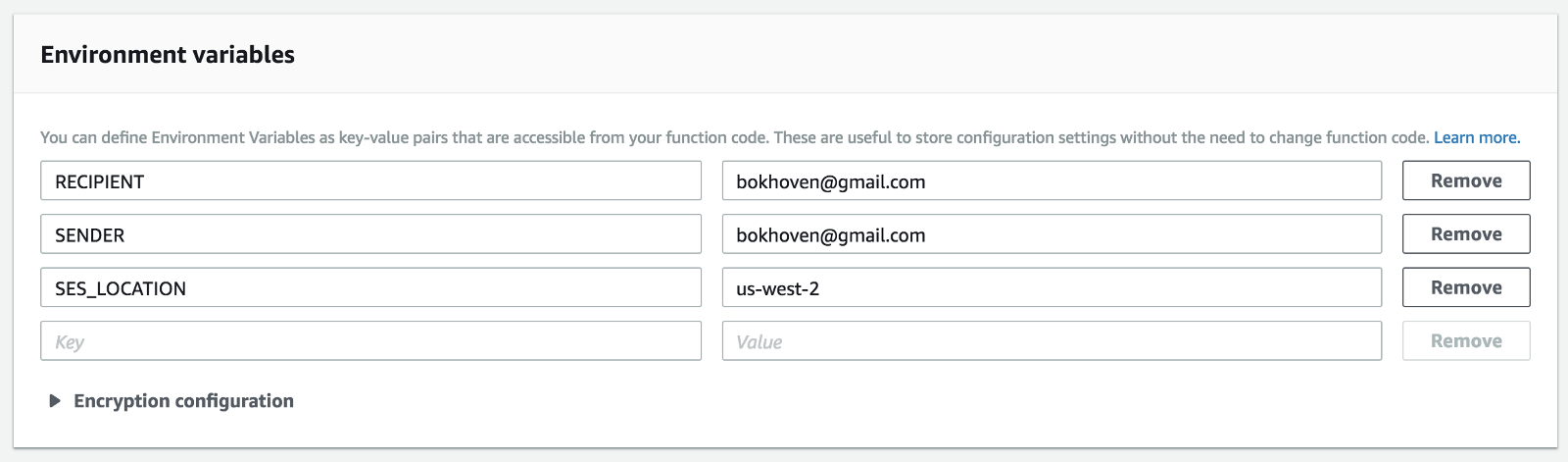
然后我们将构建Go脚本并为AWS Lambda压缩它。

使用Go运行时创建一个AWS Lambda函数，并选择或创建一个可以访问AWS SES的角色。我们将使用AWS SES发送电子邮件通知。

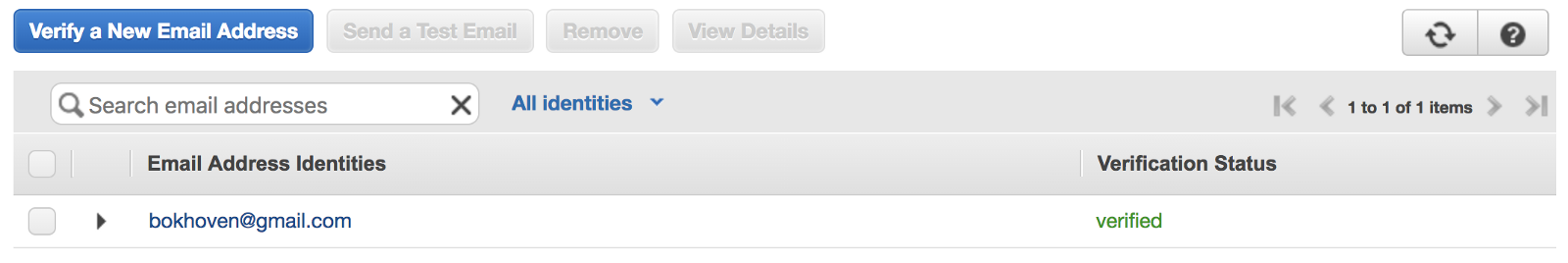
创建AWS Lambda函数后，上载zip文件并确保将处理程序设置为main。



接下来，创建3个环境变量：RECIPIENT是接收通知的电子邮件，SENDER是发送通知的电子邮件地址，last SES\_LOCATION是您的SES的位置（即：us-west-2）。

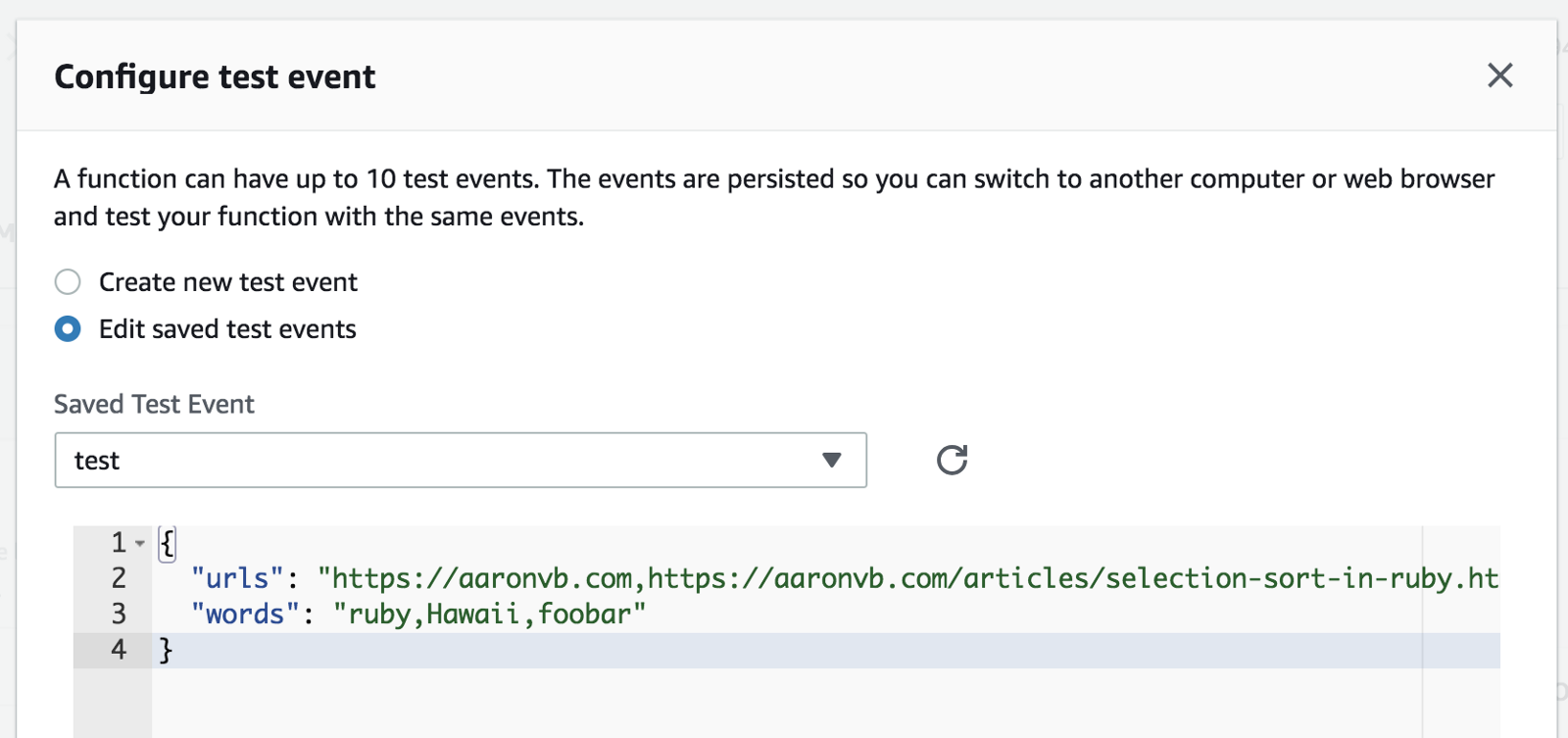


别忘了将电子邮件地址添加到SES并验证它，以便它可以接收电子邮件。



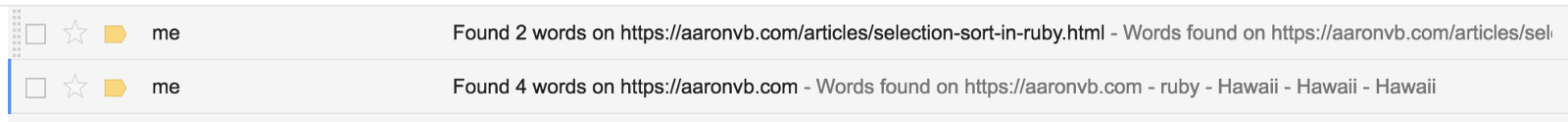
现在我们可以创建一个测试事件。在事件数据中，传递一个JSON散列，该散列有一个关键字url和一个字符串值，其中包含要擦除的url，用逗号分隔，还有一个关键字，其中字符串值为要擦除的逗号分隔的单词。

例子：





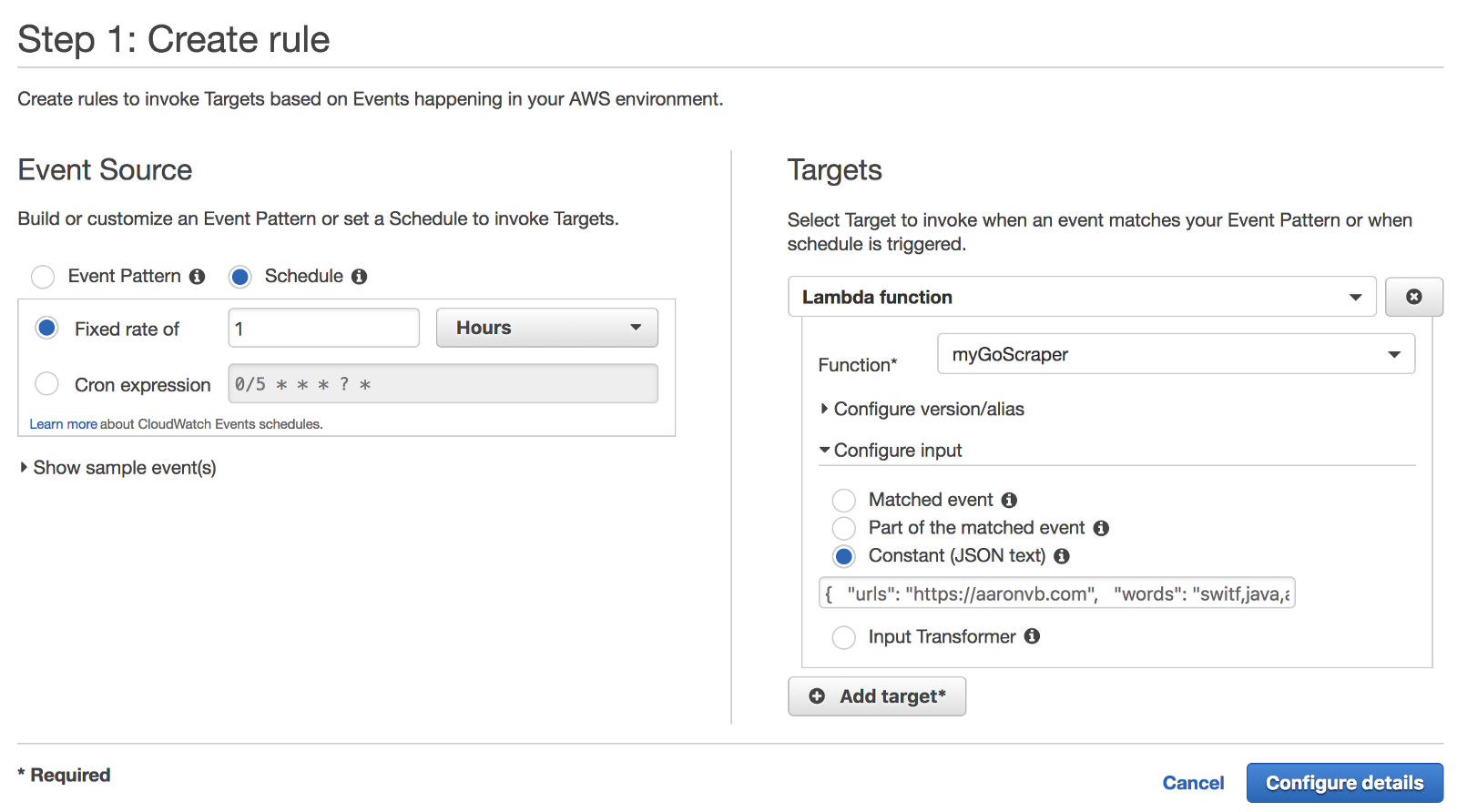
单击test按钮，您将收到一个带有日志和电子邮件的成功函数执行。日志将包含结果、来自SES的消息ID以及解析或发送电子邮件时的任何错误。



### 让我们把这个自动化

现在AWS Lambda函数正在工作，是时候自动化它并让它每小时运行一次了。我们会选择不同的词，因为我们知道它们存在。假设我们想知道我的个人网站何时会更新为“swift、java和angular”

为此，我们将使用AWS CloudWatch事件。所以让我们去那里，创建一个新的事件规则。



First we set the schedule to a fixed rate of 1 hour. Next, choose the Lambda function we created earlier. And finally, the most important part, select Configure input > Constant (JSON text), and paste in the JSON with the data to send to our function (see code below).

Once you fill that in, click Configure details to name the rule and then create it. We now have the script running every hour, scraping our website, searching for the text we provided, and alerting us when it finds it.