Using Analyzers in Elasticsearch

Aaron Rosenmund

AUTHOR EVANGELIST – SECURITY OPERATIONS

@arosenmund   www.AaronRosenmund.com

Overview

Looking under the hood:

- What Are Analyzers
- Character Filters
- Tokenizers
- Token Filters

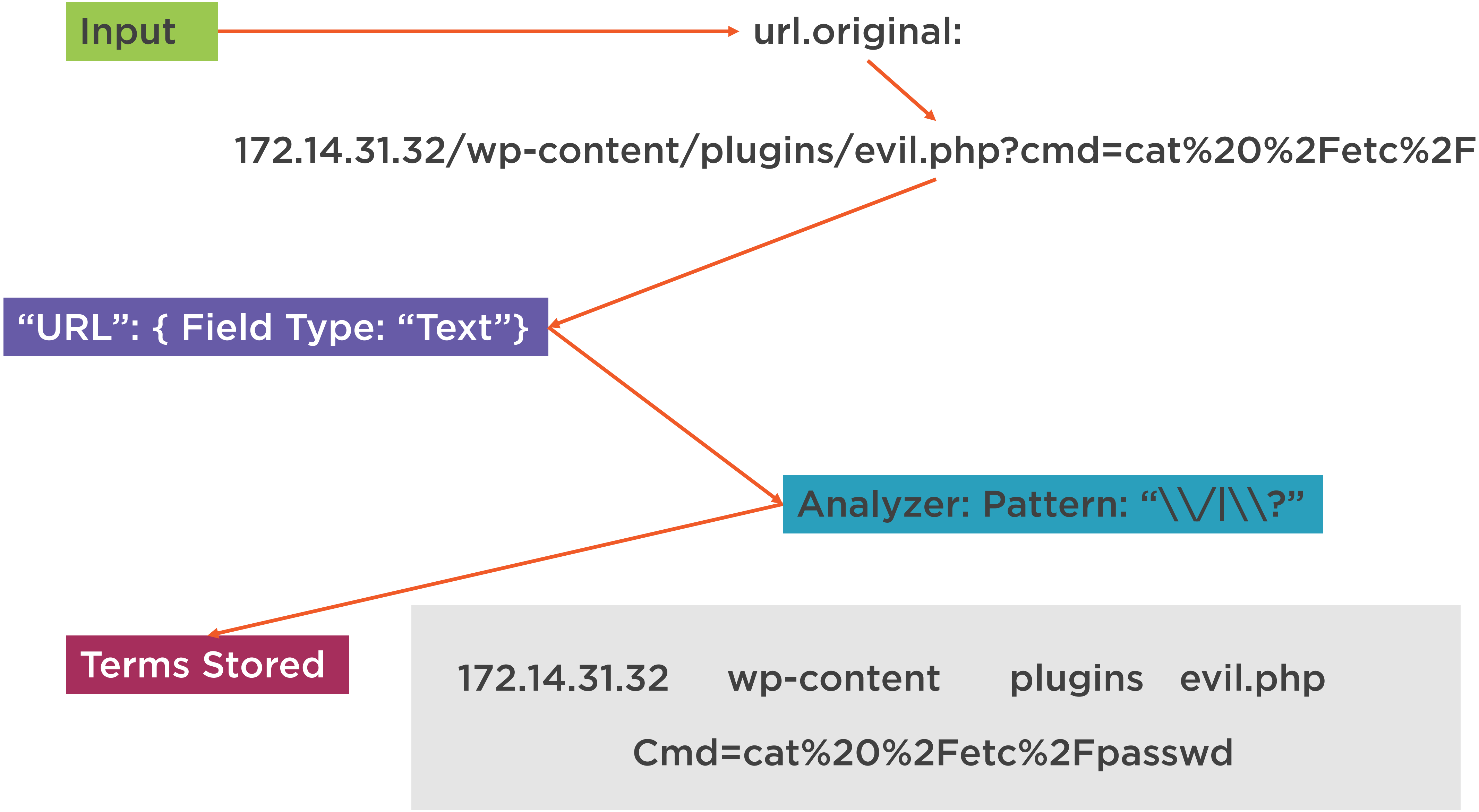
Modifying Analyzers for Search

What Is an Analyzer?

Analyzer

Algorithms made up of tokenizers, token filters, and character filters that determine how strings within a text field are transformed into terms and stored in the index.

Where Is Analysis Happening?



Input

url.original:

172.14.31.32/wp-content/plugins/evil.php?cmd=cat%20%2Fetc%2Fpasswd

“URL”: { Field Type: “Text”}

Analyzer: Pattern: “\\/|\\?”

Terms Stored

172.14.31.32 wp-content plugins evil.php

Cmd=cat%20%2Fetc%2Fpasswd

Built in Analyzers

Standard – default for all *text* fields

Simple – divides terms on any non letter character

Whitespace – divides terms on whitespace or (spaces)

Stop – supports the removal of words as defined by the user

Built in Analyzers

Keyword – outputs text as one term without modification

Pattern – filters based on regex pattern

Language – specific to a written language

Fingerprint – text cluster algorithm from the OpenRefine project

Tokenization

Breaking a text down into smaller chunks or parts, called tokens.

Custom Analyzers

tokenizer

character filter

token filter

Pre-defined combinations of tokenizers, character filters and token filters are what make up the built-in analyzer types

Analyze API

Dev tools

```
GET /_analyze

{

 "analyzer": "standard",

 "text": "172.14.31.32/wp-content/plugins/
        evil.php?cmd=cat%20%2Fetc%2Fpasswd",

 "explain": true

}
```

```
Token:”wp”


Token:”conent”
```

Demo

Demo analyze api and example data with various types of tokenization

Applying Analyzers

Creating the Analyzer

In the Index Settings

Index mapping

```
PUT securityinfo-v3
{
  “settings”: {
    “analysis”: {
      “analyzer”: {
         “url_pattern_analyzer”: {
           “type”: “pattern”,
           “pattern”: “\\/|\\?”
  }}}}
}
```

You can also specify this analyzer as the default for the index.

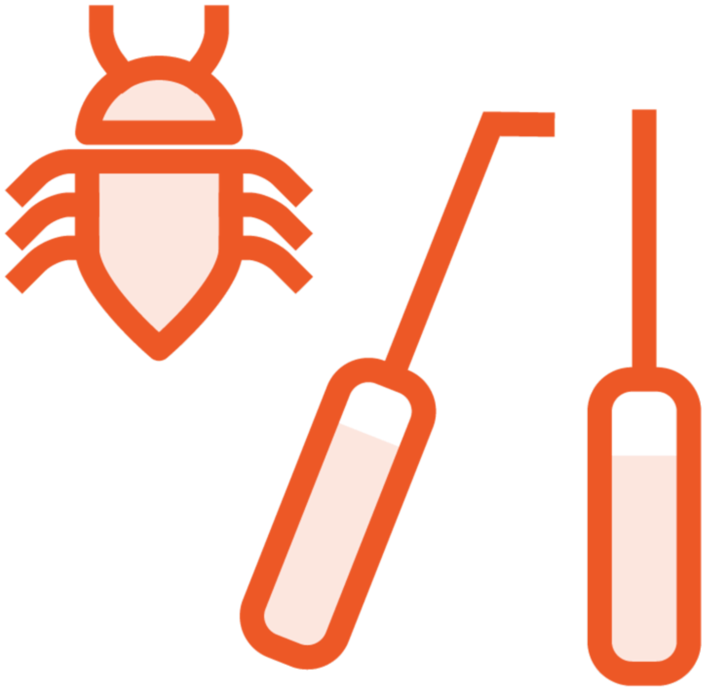
Under “analyzer” place:

```
“default”: {

  “type”:simple

}
```

Analysis by Field

Index mappings

```
PUT /securityinfo-v3/_mapping
{
 “url”: {
    properties: {
      url.original: {
        “analyzer”: “url_pattern_analyzer”
    }
}
}
```



For the intrusions yet to be detected

Summary

Applying Analyzers for customized tokenization

Resources:

Security Event Triage Path | Pluralsight

Next Step Courses:

Perform Basic Search Functions in Kibana Query Language (KQL)

Build Visualizations and Dashboards in Kibana