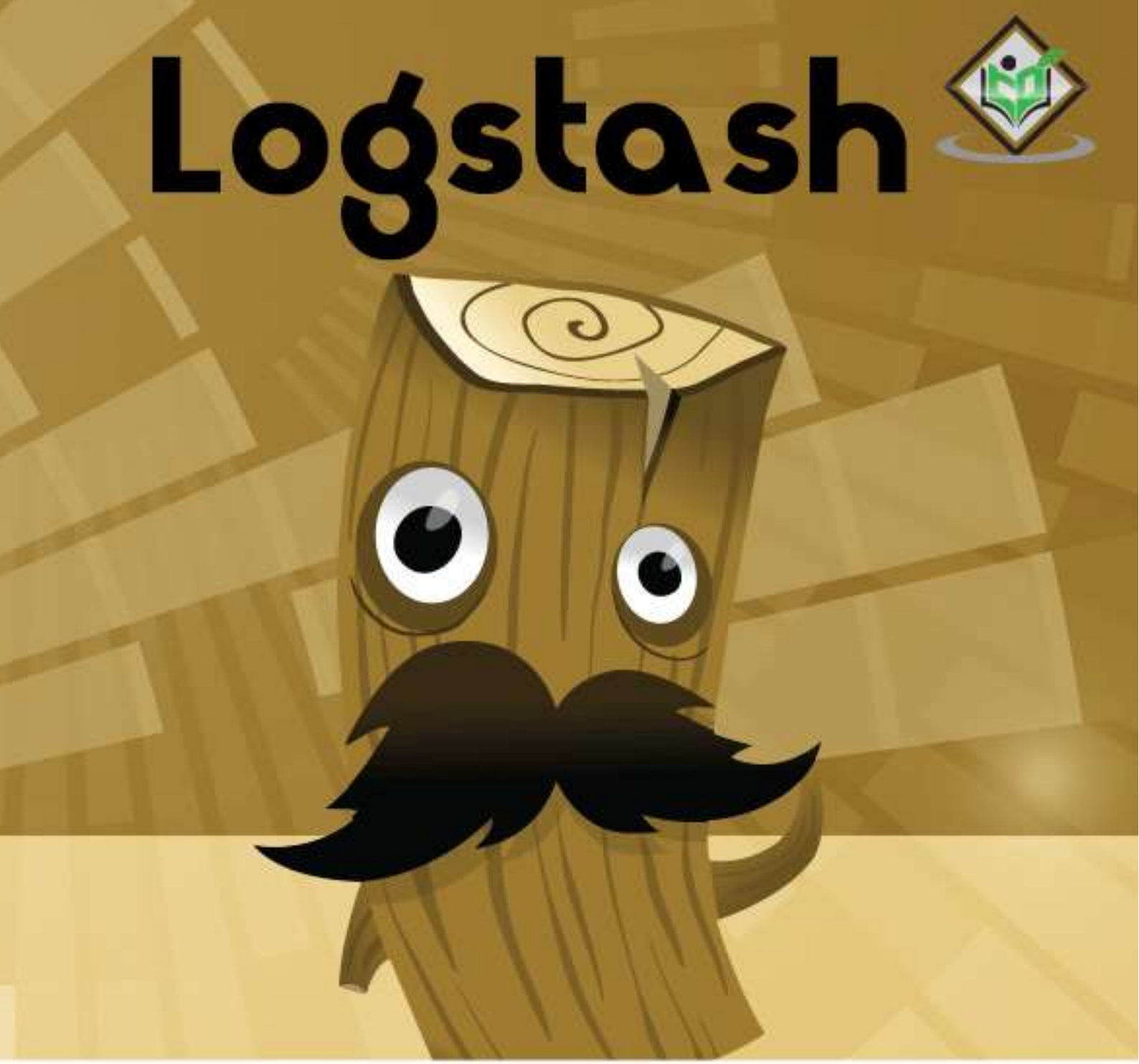
Logstash



tutorialspoint

SIMPLY EASY LEARNING

www.tutorialspoint.com

https://www.facebook.com/tutorialspointindia https://twitter.com/tutorialspoint

About the Tutorial

Logstash is an open-source, centralized, events and logging manager. It is a part of the ELK (ElasticSearch, Logstash, Kibana) stack. In this tutorial, we will understand the basics of Logstash, its features, and the various components it has.

Audience

This tutorial is designed for software professionals who want to learn the basics of Logstash and its programming concepts in simple and easy steps. It describes the components and functions of Logstash with suitable examples.

Prerequisites

The readers are expected to have a basic understanding of Ruby, JSON, and web technologies. Additionally it will be helpful for the readers to be familiar with Logging Techniques and Regex patterns.

Copyright and Disclaimer

Table of Contents

1. Logstash – Introduction

Logstash is a tool based on the filter/pipes patterns for gathering, processing and generating the logs or events. It helps in centralizing and making real time analysis of logs and events from different sources.

Logstash is written on JRuby programming language that runs on the JVM, hence you can run Logstash on different platforms. It collects different types of data like Logs, Packets, Events, Transactions, Timestamp Data, etc., from almost every type of source. The data source can be Social data, E-commerce, News articles, CRM, Game data, Web trends, Financial data, Internet of Things, Mobile devices, etc.

Logstash – General Features

The general features of Logstash are as follows:

- Logstash can collect data from different sources and send to multiple destinations.
- Logstash can handle all types of logging data like Apache Logs, Windows Event Logs, Data over Network Protocols, Data from Standard Input and many more.
- Logstash can also handle http requests and response data.
- Logstash provides a variety of filters, which helps the user to find more meaning in the data by parsing and transforming it.
- Logstash can also be used for handling sensors data in internet of things.
- Logstash is open source and available under the Apache license version 2.0.

Logstash – Key Concepts

The key concepts of Logstash are as follows:

Event Object

It is the main object in Logstash, which encapsulates the data flow in the Logstash pipeline. Logstash uses this object to store the input data and add extra fields created during the filter stage.

Logstash offers an Event API to developers to manipulate events. In this tutorial, this event is referred with various names like Logging Data Event, Log Event, Log Data, Input Log Data, Output Log Data, etc.

Pipeline

It comprises of data flow stages in Logstash from input to output. The input data is entered in the pipeline and is processed in the form of an event. Then sends to an output destination in the user or end system's desirable format.

Input

This is the first stage in the Logstash pipeline, which is used to get the data in Logstash for further processing. Logstash offers various plugins to get data from different platforms. Some of the most commonly used plugins are – File, Syslog, Redis and Beats.

Filter

This is the middle stage of Logstash, where the actual processing of events take place. A developer can use pre-defined Regex Patterns by Logstash to create sequences for differentiating between the fields in the events and criteria for accepted input events.

Logstash offers various plugins to help the developer to parse and transform the events into a desirable structure. Some of the most commonly used filter plugins are – Grok, Mutate, Drop, Clone and Geoip.

Output

This is the last stage in the Logstash pipeline, where the output events can be formatted into the structure required by the destination systems. Lastly, it sends the output event after complete processing to the destination by using plugins. Some of the most commonly used plugins are – Elasticsearch, File, Graphite, Statsd, etc.

Logstash – Advantages

The following points explain the various advantages of Logstash.

- Logstash offers regex pattern sequences to identify and parse the various fields in any input event.
- Logstash supports a variety of web servers and data sources for extracting logging data.
- Logstash provides multiple plugins to parse and transform the logging data into any user desirable format.
- Logstash is centralized, which makes it easy to process and collect data from different servers.
- Logstash supports many databases, network protocols and other services as a destination source for the logging events.
- Logstash uses the HTTP protocol, which enables the user to upgrade Elasticsearch versions without having to upgrade Logstash in a lock step.

Logstash – Disadvantages

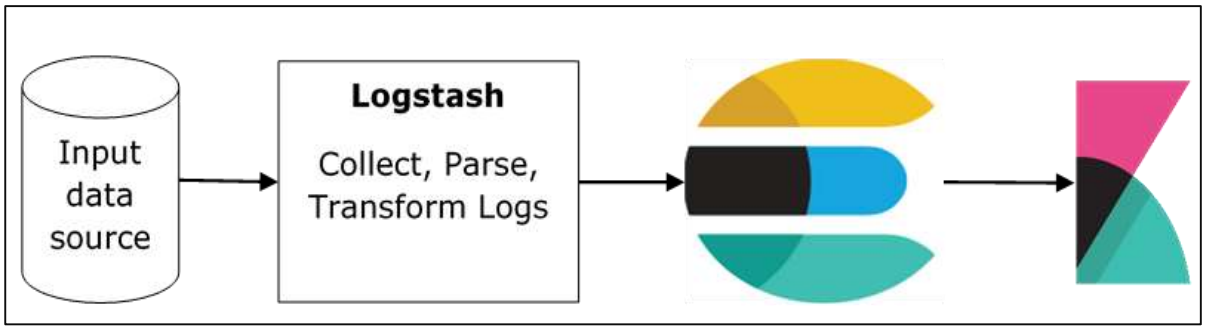
The following points explain the various disadvantages of Logstash.

- Logstash uses http, which negatively affects the processing of the logging data.
- Working with Logstash can sometimes be a little complex, as it needs a good understanding and analysis of the input logging data.
- Filter plugins are not generic, so, the user may need to find the correct sequence of patterns to avoid error in parsing.

In the next chapter, we will understand what the ELK Stack is and how it helps Logstash.

2. Logstash – ELK Stack

ELK stands for **Elasticsearch, Logstash,** and **Kibana**. In the ELK stack, Logstash extracts the logging data or other events from different input sources. It processes the events and later stores it in Elasticsearch. Kibana is a web interface, which accesses the logging data form Elasticsearch and visualizes it.



Logstash and Elasticsearch

Logstash provides input and output Elasticsearch plugin to read and write log events to Elasticsearch. Elasticsearch as an output destination is also recommended by Elasticsearch Company because of its compatibility with Kibana. Logstash sends the data to Elasticsearch over the http protocol.

Elasticsearch provides bulk upload facility, which helps to upload the data from different sources or Logstash instances to a centralized Elasticsearch engine. ELK has the following advantages over other DevOps Solutions:

- ELK stack is easier to manage and can be scaled for handling petabytes of events.
- ELK stack architecture is very flexible and it provides integration with Hadoop. Hadoop is mainly used for archive purposes. Logstash can be directly connected to Hadoop by using flume and Elasticsearch provides a connector named **es-hadoop** to connect with Hadoop.
- ELK ownership total cost is much lesser than its alternatives.

Logstash and Kibana

Kibana does not interact with Logstash directly but through a data source, which is Elasticsearch in the ELK stack. Logstash collects the data from every source and Elasticsearch analyzes it at a very fast speed, then Kibana provides the actionable insights on that data.

Kibana is a web based visualization tool, which helps developers and others to analyze the variations in large amounts of events collected by Logstash in Elasticsearch engine. This visualization makes it easy to predict or to see the changes in trends of errors or other significant events of the input source.

3. Logstash – Installation

To install Logstash on the system, we should follow the steps given below:

Step 1: Check the version of your Java installed in your computer; it should be Java 8 because it is not compatible with Java 9. You can check this by –

In a Windows Operating System (OS) (using command prompt):

```
> java -version
```

In UNIX OS (Using Terminal):

```
$ echo $JAVA_HOME
```

Step 2: Download Logstash from – https://www.elastic.co/downloads/logstash.

- For Windows OS, download the ZIP file.
- For UNIX OS, download the TAR file.
- For Debian OS download the DEB file.
- For Red Hat and other Linux distributions, download the RPN file.
- APT and Yum utilities can also be used to install Logstash in many Linux distributions.

Step 3: The installation process for Logstash is very easy. Let's see how you can install Logstash on different platforms.

Note: Do not put any whitespace or colon in the installation folder.

- **Windows OS:** Unzip the zip package and the Logstash is installed.
- **UNIX OS:** Extract the tar file in any location and the Logstash is installed.

```
$tar –xvf logstash-5.0.2.tar.gz
```

- **Using APT utility for Linux OS:**
  - Download and install the Public Signing Key:

```
$ wget -qO - https://artifacts.elastic.co/GPG-KEY-elasticsearch | sudo
apt-key add -
```

- Save the repository definition:

```
$ echo "deb https://artifacts.elastic.co/packages/5.x/apt stable main" |
sudo tee -a /etc/apt/sources.list.d/elastic-5.x.list
```

- Run update:

```
$ sudo apt-get update
```

- Now you can install by using the following command:

```
$ sudo apt-get install logstash
```

- **Using YUM utility for Debian Linux OS**:
  - Download and install the Public Signing Key:

```
$ rpm --import https://artifacts.elastic.co/GPG-KEY-elasticsearch
```

  - Add the following text in the file with the .repo suffix in your "/etc/yum.repos.d/" directory. For example, **logstash.repo**

```
[logstash-5.x]

name=Elastic repository for 5.x packages

baseurl=https://artifacts.elastic.co/packages/5.x/yum

gpgcheck=1

gpgkey=https://artifacts.elastic.co/GPG-KEY-elasticsearch

enabled=1

autorefresh=1

type=rpm-md
```

  - You can now install Logstash by using the following command:

```
$ sudo yum install logstash
```

Step 4: Go to the Logstash home directory. Inside the bin folder, run the **elasticsearch.bat** file in case of windows or you can do the same using the command prompt and through the terminal. In UNIX, run the Logstash file.

We need to specify the input source, output source and optional filters. For verifying the installation, you can run it with the basic configuration by using a standard input stream (stdin) as the input source and a standard output stream (stdout) as the output source. You can specify the configuration in the command line also by using **–e** option.

In Windows:

```
> cd logstash-5.0.1/bin

> Logstash -e 'input { stdin { } } output { stdout {} }'
```

In Linux:

```
$ cd logstash-5.0.1/bin

$ ./logstash -e 'input { stdin { } } output { stdout {} }'
```

Note: in case of windows, you might get an error stating JAVA_HOME is not set. For this, please set it in environment variables to "C:\Program Files\Java\jre1.8.0_111" or the location where you installed java.

Step 5: Default ports for Logstash web interface are 9600 to 9700 are defined in the **logstash-5.0.1\config\logstash.yml** as the **http.port** and it will pick up the first available port in the given range.

We can check if the Logstash server is up and running by browsing http://localhost:9600 or if the port is different and then please check the command prompt or terminal. We can see the assigned port as "Successfully started Logstash API endpoint {:port=>9600}. It will return a JSON object, which contains the information about the installed Logstash in the following way:

```
{

"host":"manu-PC",

"version":"5.0.1",

"http_address":"127.0.0.1:9600",

"build_date":"2016-11-11T22:28:04+00:00",

"build_sha":"2d8d6263dd09417793f2a0c6d5ee702063b5fada",

"build_snapshot":false

}
```