Onstak, Inc.

Corporate Headquarters

1525 McCarthy Blvd, Suite 1123,

Milpitas, CA 95035, USA

+1-408-834-4869

[www.onstak.com](http://www.onstak.com)

[info@onstak.com](mailto:info@onstak.com)

**Splunk and Kafka Documentation**

[1. Splunk Installation on CentOS 7 3](#_Toc1016275340)

[2. Splunk Connect for Kafka Installation 4](#_Toc665764199)

[2.1 System Requirements 4](#_Toc1517030818)

[2.2 Configuring Splunk 4](#_Toc2112842732)

[3. Connector Installation and Configuration 10](#_Toc1742346522)

[4. References 14](#_Toc1095011631)

# Splunk Installation on CentOS 7

1. Login with credentials of your machine and then become root user by following command.  
   sudo su -
2. Create your account on Splunk.com and download your Splunk package in my case i used the following command to download.  
   wget -O splunk-8.0.1-6db836e2fb9e-linux-2.6-x86\_64.rpm 'https://www.splunk.com/bin/splunk/DownloadActivityServlet?architecture=x86\_64&platform=linux&version=8.0.1&product=splunk&filename=splunk-8.0.1-6db836e2fb9e-linux-2.6-x86\_64.rpm&wget=true'
3. Install Splunk   
   rpm -i (Splunk-version-rpm-package)  
   in my case rpm -i splunk-8.0.1-6db836e2fb9e-linux-2.6-x86\_64.rpm
4. After installation is complete you need to create the admin credentials manually without starting Splunk.   
   Open the user.conf file with your favourite editor. I used vi  
   vi/opt/splunk/etc/system/local/user-seed.conf  
   Now add the following lines in file  
   [user\_info]  
   USERNAME = admin   
   PASSWORD = your password
5. Start your Splunk with following command  
   /opt/splunk/bin/splunk start
6. From your web browser open Splunk Enterprise Web Console   
   <http://your_public_ip:8000>Enter your credentials and explore Splunk.

Your Splunk machine is now ready to go.

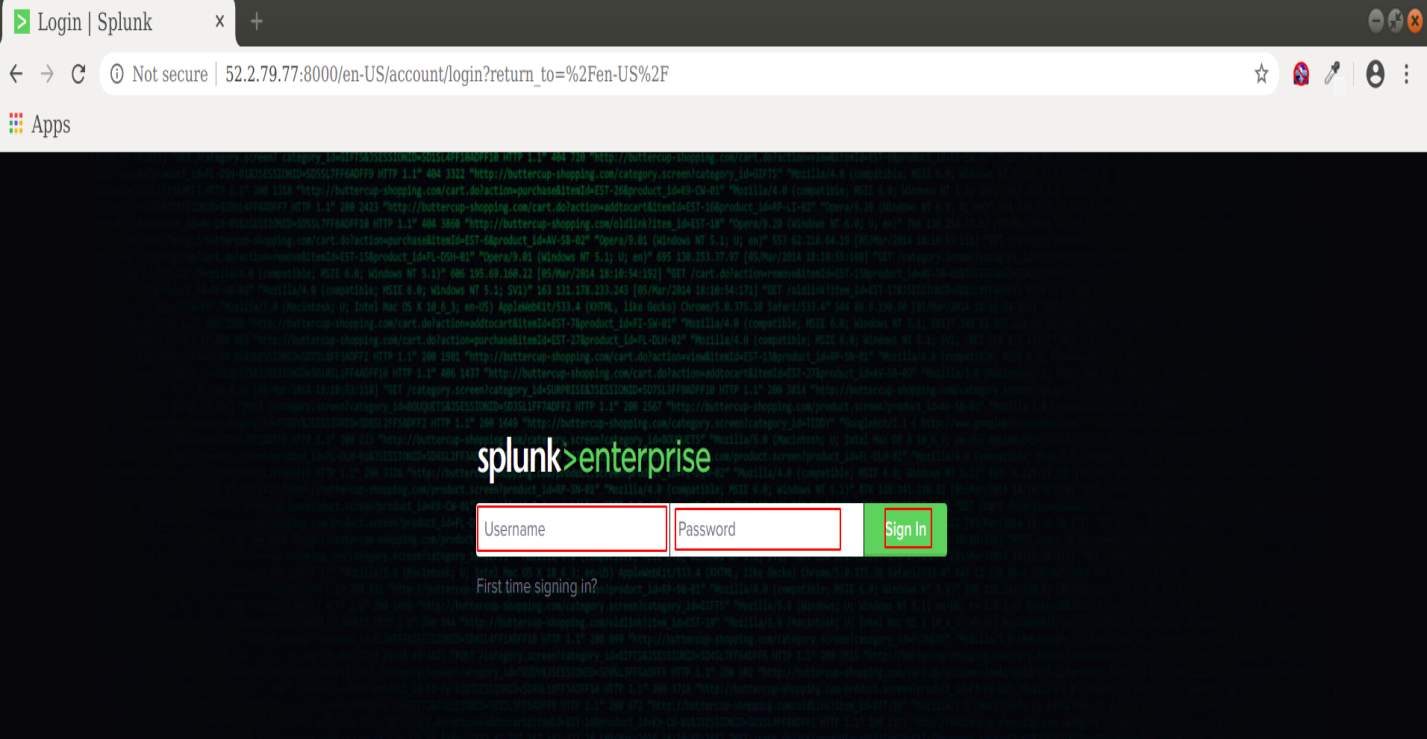
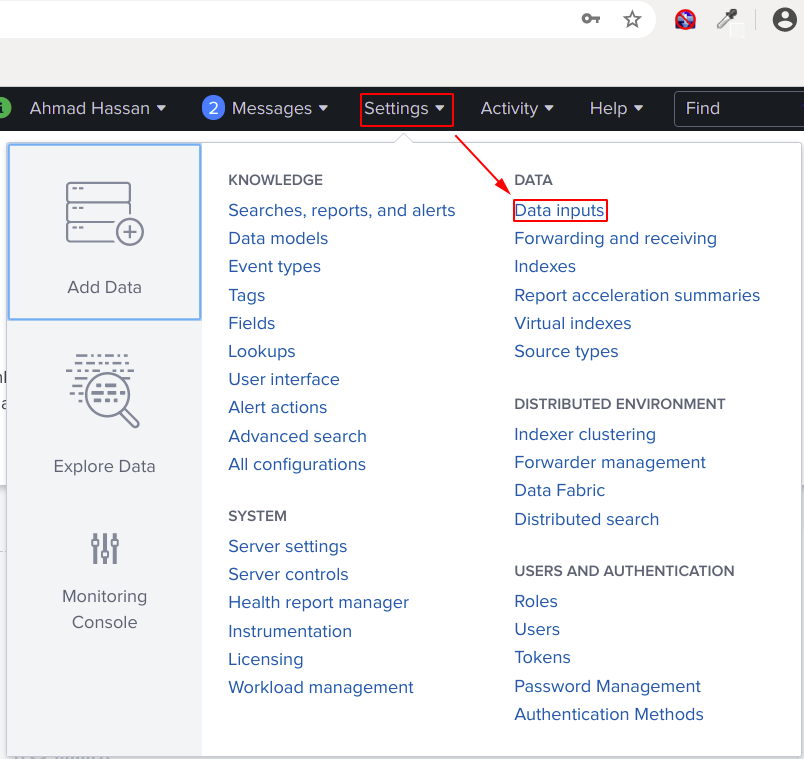
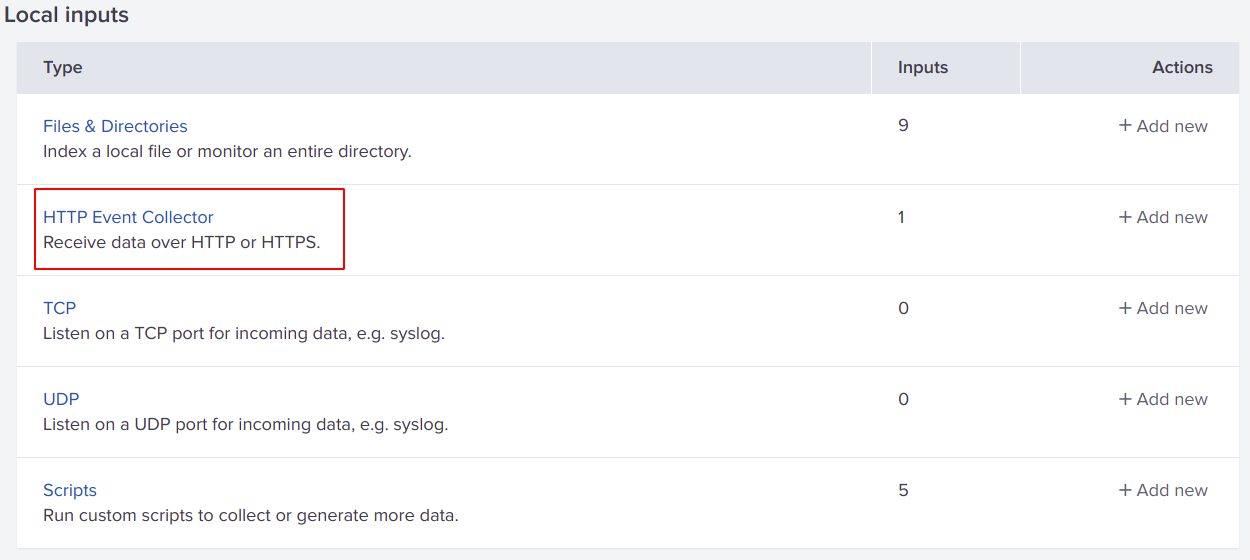
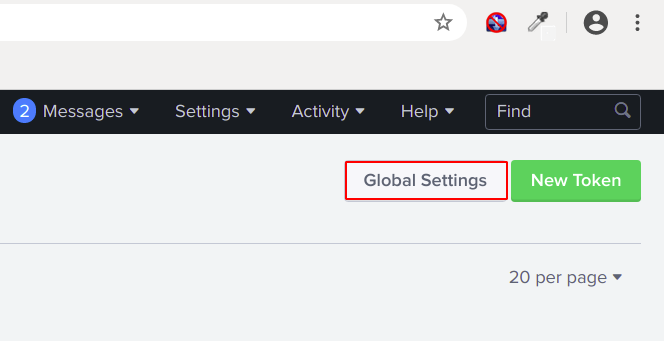
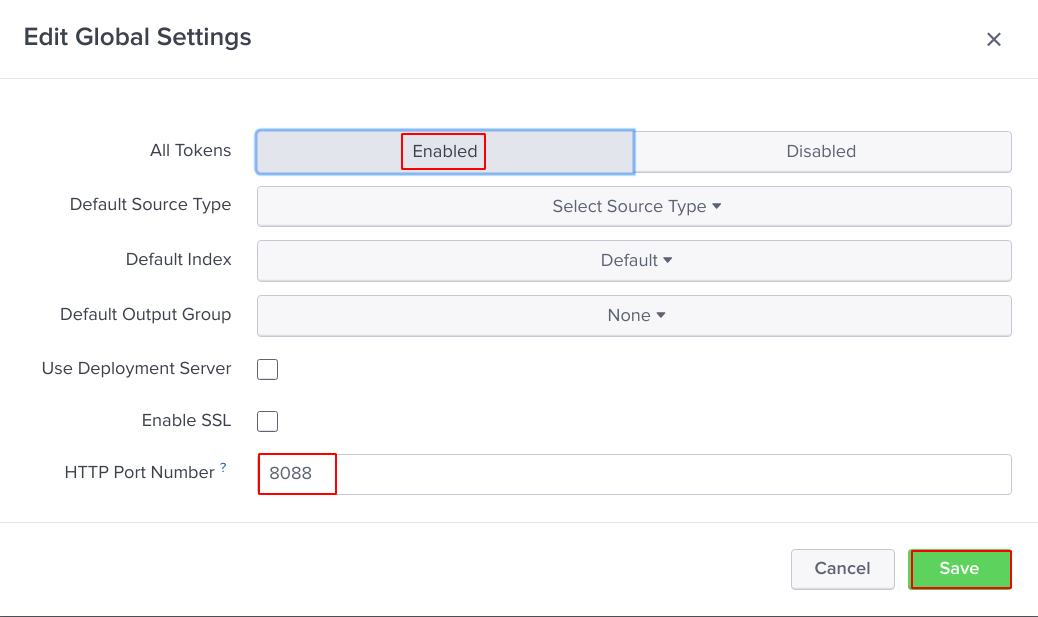
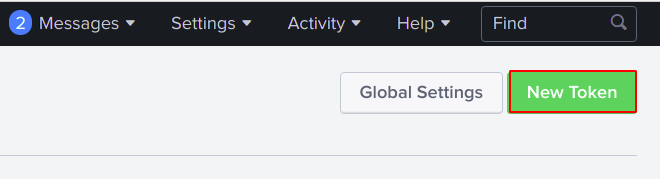
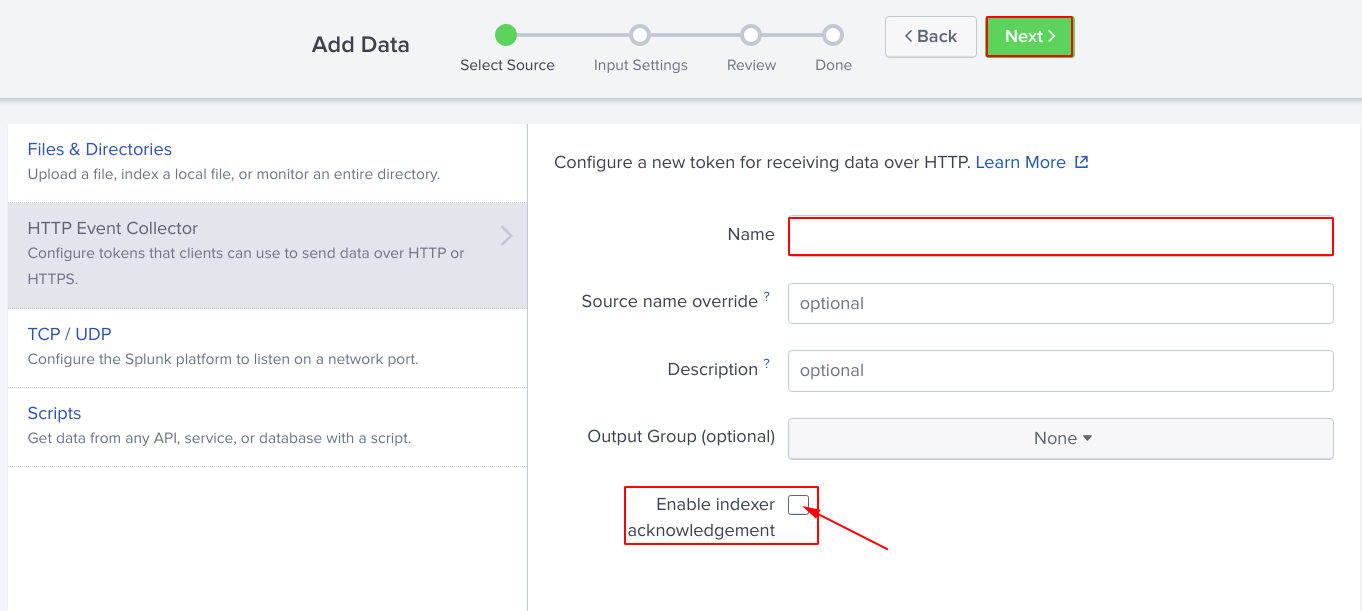
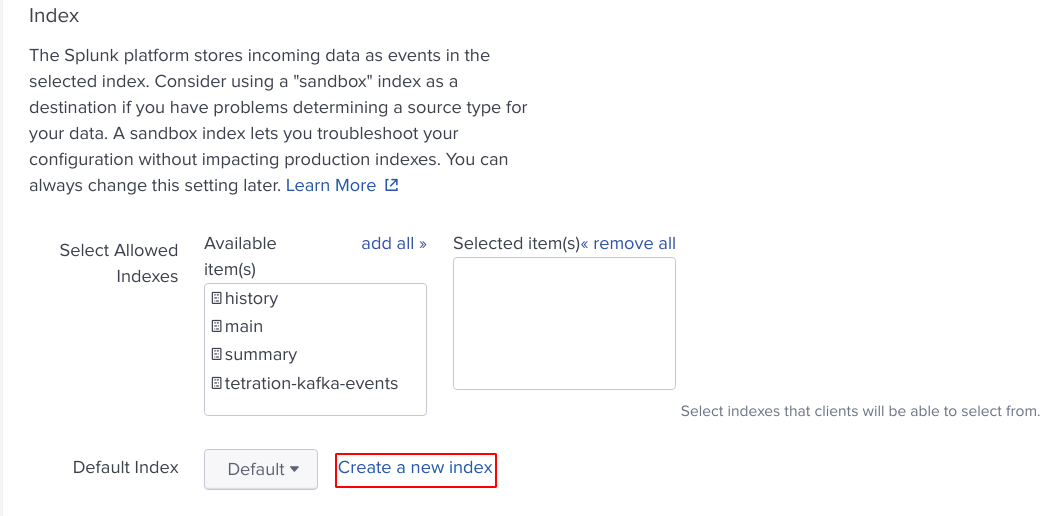
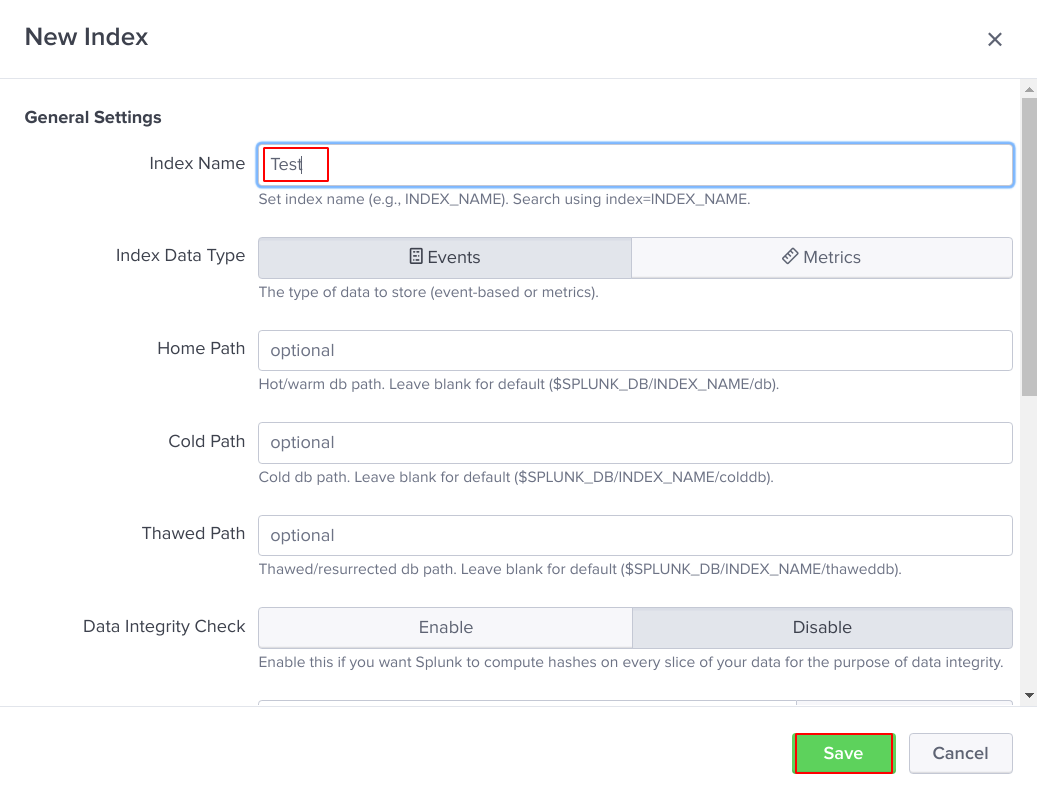
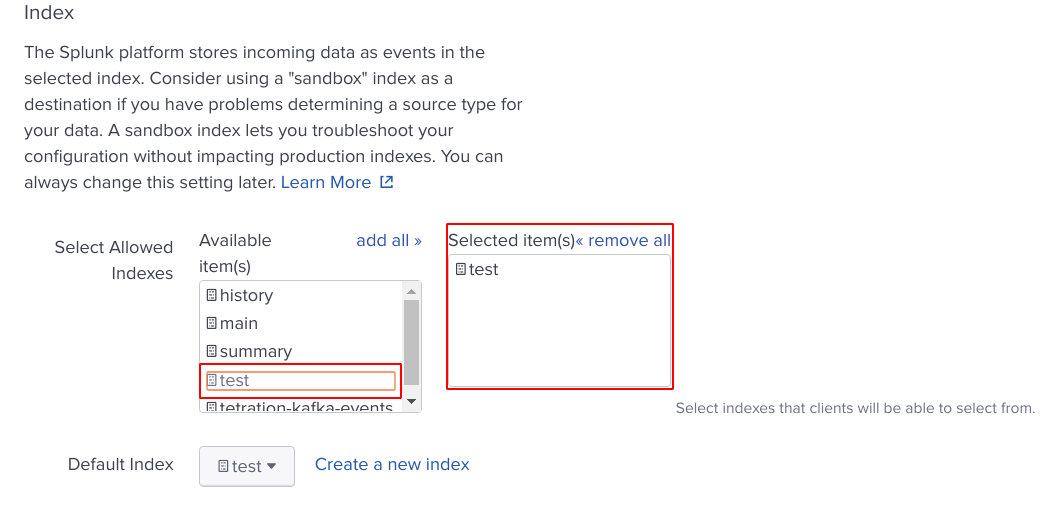
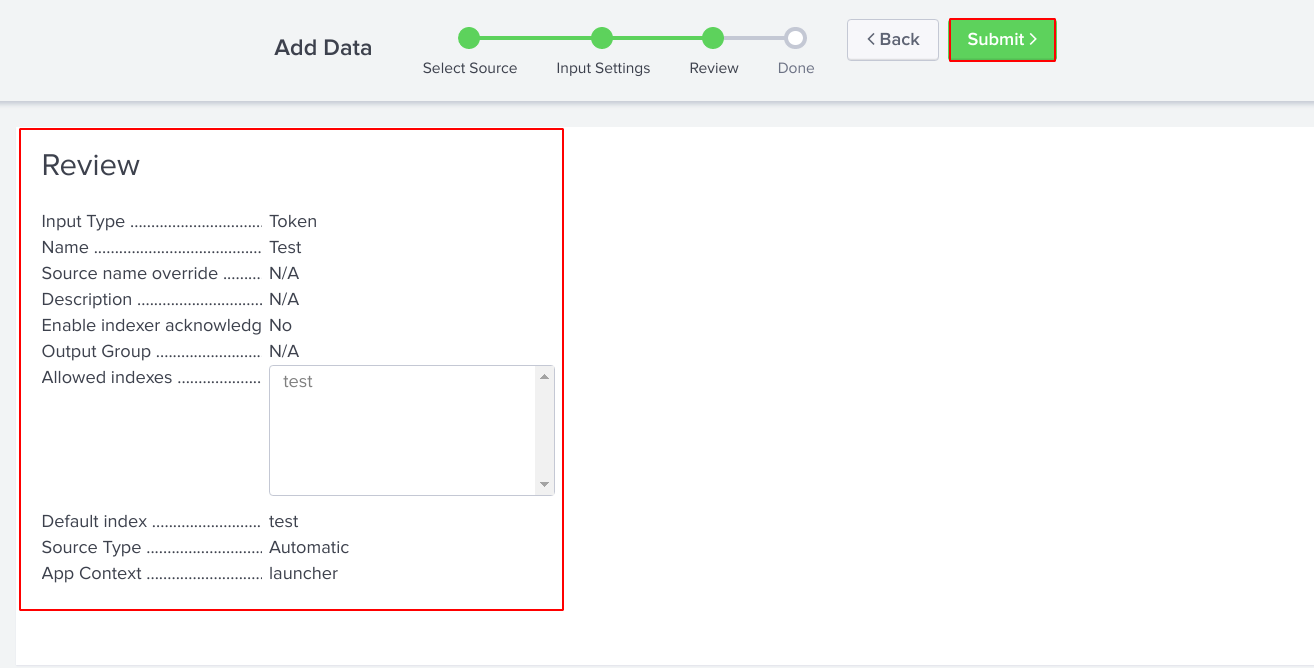
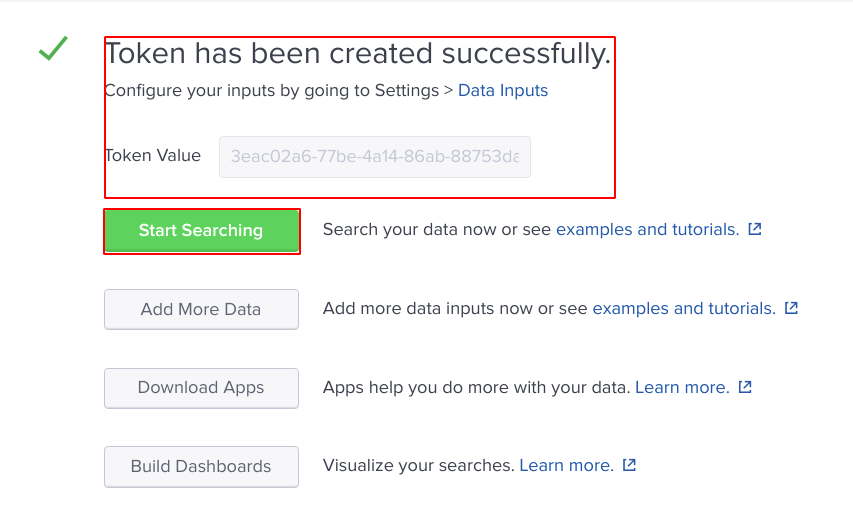
# Splunk Connect for Kafka Installation

## System Requirements

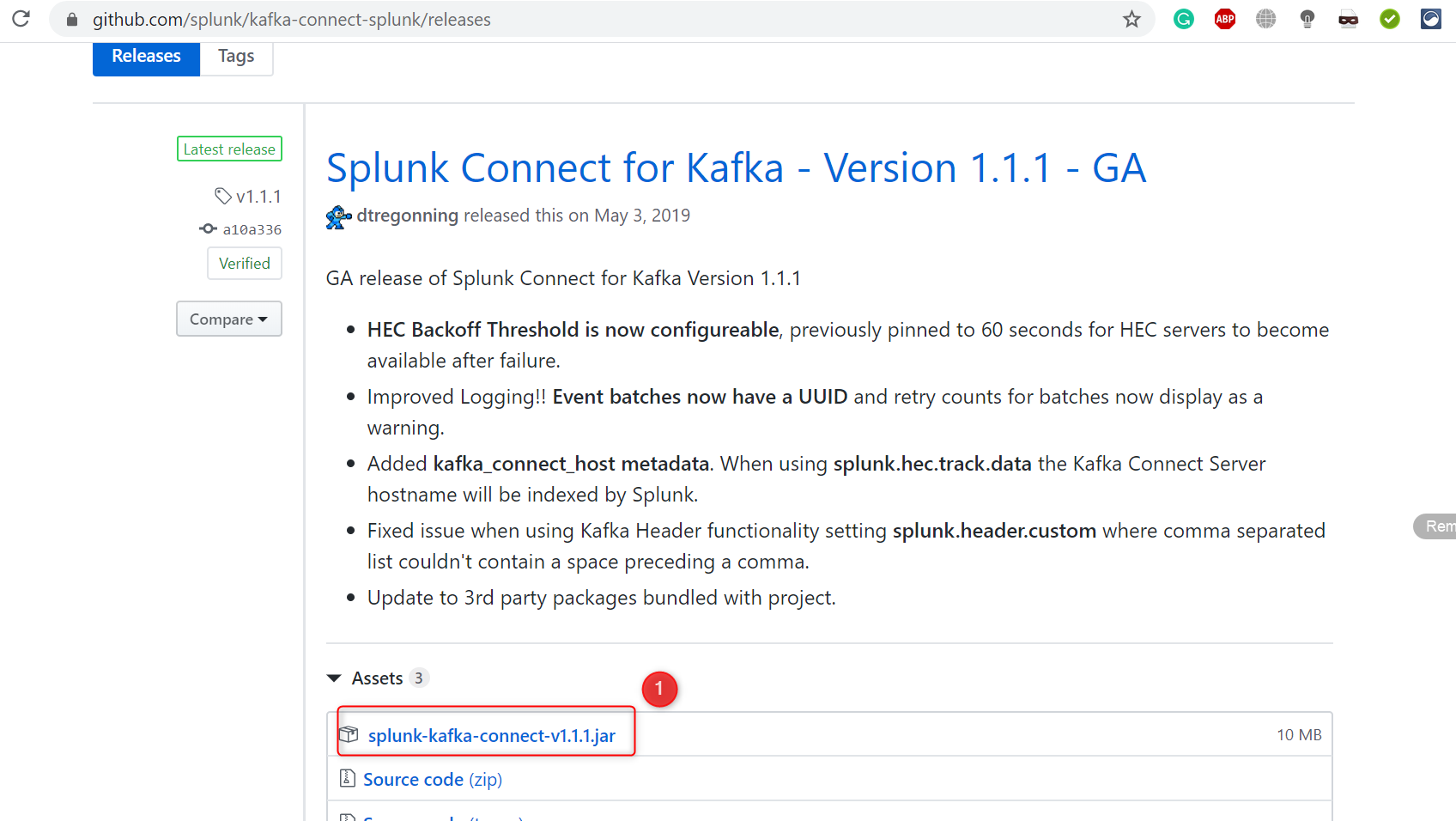
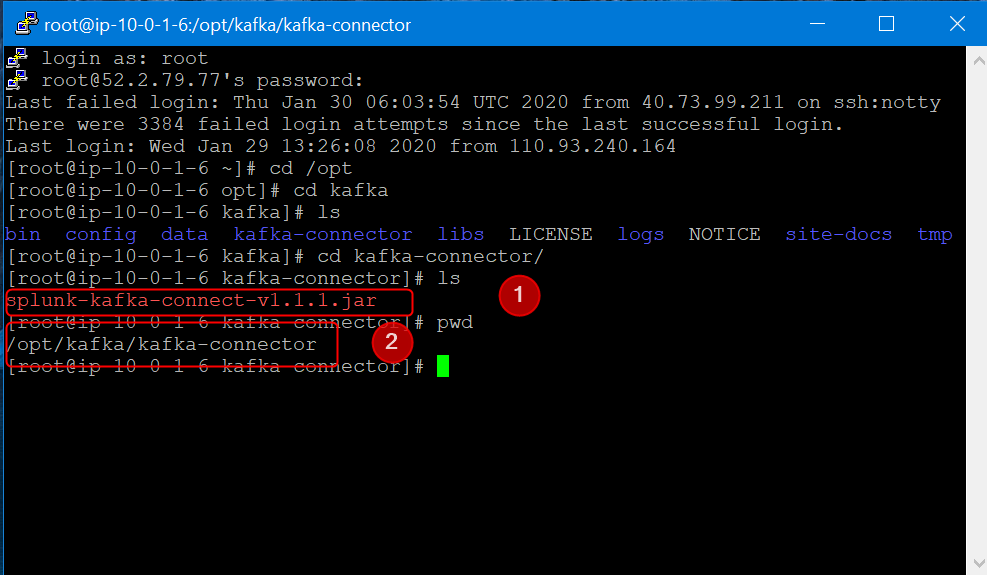
A Kafka Connect Environment running Kafka version 1.0.0 or later.  
 Java 8 or later.  
 Splunk platform environment of version 6.5 or later.  
 Configured and Valid HTTP Event Collector (HEC) token.

## Configuring Splunk

Start creating HEC token by following the instructions provided below with images.

1. From any browser visit <http://your_public_ip:8000> and provide your credentials to login.  
     
   
2. After login click **Settings**  and under setting click **Data inputs** as shown in below image.  
     
   
3. Next you will click on HTTP Event Collector  
   
4. Before start creating New Token click **Global Setting**
5. Here you will **Enabled**  all tokens and verify the port **8088.** Then **Save** the settings.
6. Now lets create our **New Token.**
7. Give your token any **Name** and according to your need you can enable or disable **Indexer acknowledgement** and then click **Next.**
8. Lets **Create a new index** or you can use any existing.  
   
9. Simple add a **Name** for your index and click **Save.** You can leave other options as default.  
   
10. Select your created index and make sure it appears in **Selected item(s)** box and click **Review.**
11. Review your setting and click **Submit** to create your token.  
      
    
12. Here your token is created you can copy your **Token Value** and can use in your Rest Api.  
    And you can start searching now.  
    

# Connector Installation and Configuration

1. Visit <https://github.com/splunk/kafka-connect-splunk/releases> and download the latest **splunk-kafka-connect-[version].jar** release.  
   
2. Create a directory to store your Kafka Connect Connector. This will be used for your **plugin.path** setting.  
   
3. Navigate to your /$KAFKA\_HOME/config/ directory.  
   Modify the connect-distributed.properties file to include the below information.

#These settings may already be configured if you have deployed a connector in your Kafka Connect Environment

bootstrap.servers=<BOOTSTRAP\_SERVERS>(ip-10-0-1-6.ec2.internal:9092)

plugin.path=<PLUGIN\_PATH>(/opt/kafka/kafka-connector)

#Required

key.converter=org.apache.kafka.connect.storage.StringConverter

value.converter=org.apache.kafka.connect.storage.StringConverter

key.converter.schemas.enable=false

value.converter.schemas.enable=false

internal.key.converter=org.apache.kafka.connect.json.JsonConverter

internal.value.converter=org.apache.kafka.connect.json.JsonConverter

internal.key.converter.schemas.enable=false

internal.value.converter.schemas.enable=false

offset.flush.interval.ms=10000

#Recommended

group.id=kafka-connect-splunk-hec-sink

**bootstrap.servers** – This is a comma-separated list of where your Kafka brokers are located.

**plugin.path** – To make the JAR visible to Kafka Connect, we need to ensure that when Kafka Connect is started that the plugin path variable is folder path location of where your connector was installed to in the earlier section.

1. After we have our properties file configured and ready to go, we're now ready to deploy Kafka Connect. If we kept the same name for our properties file, this command will deploy Kafka Connect  
   ./bin/connect-distributed.sh config/connect-distributed.properties
2. Now is a good time to check that the Splunk Connect for Kafka has been installed correctly and is ready to be deployed. Run the following command and note the results.  
   Curl <http://KAFKA_CONNECT_HOST:8083/connector-plugins>
3. Initiate Connector.  
   Splunk Indexing with Acknowledgement Using HEC/Event Endpoints:  
   Rest API:  
     
   curl -s 10.0.1.6:8083/connectors -X POST -H “Content-Type:application/json” -d’{  
   “name”: “Sending-Data”,   
   “config”: {   
   “connector.class”: “com.splunk.kafka.connect.SplunkSinkConnector”,  
   “task.max”: “1”,  
   “topics”: “test”,  
   “splunk.indexes”: “test”,  
   “splunk.sourcetypes”: “access\_combined”,  
   “splunk.hec.uri”: “http://54.224.99.155:8088”,  
   “splunk.hec.token”: “c8926f55-2c9b-4e16-88a7-eba0a039998b”,  
   “splunk.hec.ack.enabled”: “false”,  
   “splunk.hec.raw”: “true”,  
   “splunk.hec.raw.line.breaker”: “####”,  
   “splunk.hec.ssl.validate.certs”: “false”  
    }  
   }’

Following is the information about the REST API.

1. **name**: Connector name. A consumer group with this name will be created with tasks to be distributed evenly across the connector cluster nodes.
2. **connector.class**: The Java class used to perform connector jobs. Keep the default unless you modify the connector.
3. **tasks.max**: The number of tasks generated to handle data collection jobs in parallel. The tasks will be spread evenly across all Splunk Kafka Connector nodes
4. **topics:**Comma separated list of Kafka topics for Splunk to consume
5. **splunk.hec.uri:** Splunk HEC URIs. Either a comma separated list of the FQDNs or IPs of all Splunk indexers, or a load balancer. If using the former, the connector will load balance to indexers using round robin
6. **splunk.hec.token:** Splunk HTTP Event Collector token
7. **splunk.hec.ack.enabled**: Valid settings are true or false. When set to true the Splunk Kafka Connector will poll event ACKs for POST events before check-pointing the Kafka offsets. This is used to prevent data loss, as this setting implements guaranteed delivery
8. **splunk.hec.raw**: Set to true in order for Splunk software to ingest data using the the /raw HEC endpoint. false will use the /event endpoint
9. **splunk.hec.json.event.enrichment:** Only applicable to /event HEC endpoint. This setting is used to enrich raw data with extra metadata fields. It contains a comma separated list of key value pairs. The configured enrichment metadata will be indexed along with raw event data by Splunk software. Note: Data enrichment for /event HEC endpoint is only available in Splunk Enterprise 6.5 and above
10. **splunk.hec.track.data**: Valid settings are true or false. When set to true, data loss and data injection latency metadata will be indexed along with raw data

# References

<https://www.digitalocean.com/community/tutorials/how-to-install-apache-kafka-on-centos-7>

<https://linuxconfig.org/how-to-install-kafka-on-redhat-8>

<https://docs.splunk.com/Documentation/KafkaConnect/1.1.1/User/InstallSplunkKafkaConnect>

<https://www.splunk.com/en_us/blog/tips-and-tricks/splunk-connect-for-kafka-connecting-apache-kafka-with-splunk.html>

<https://blog.redbranch.net/2018/04/19/zookeeper-install-on-centos-7/>