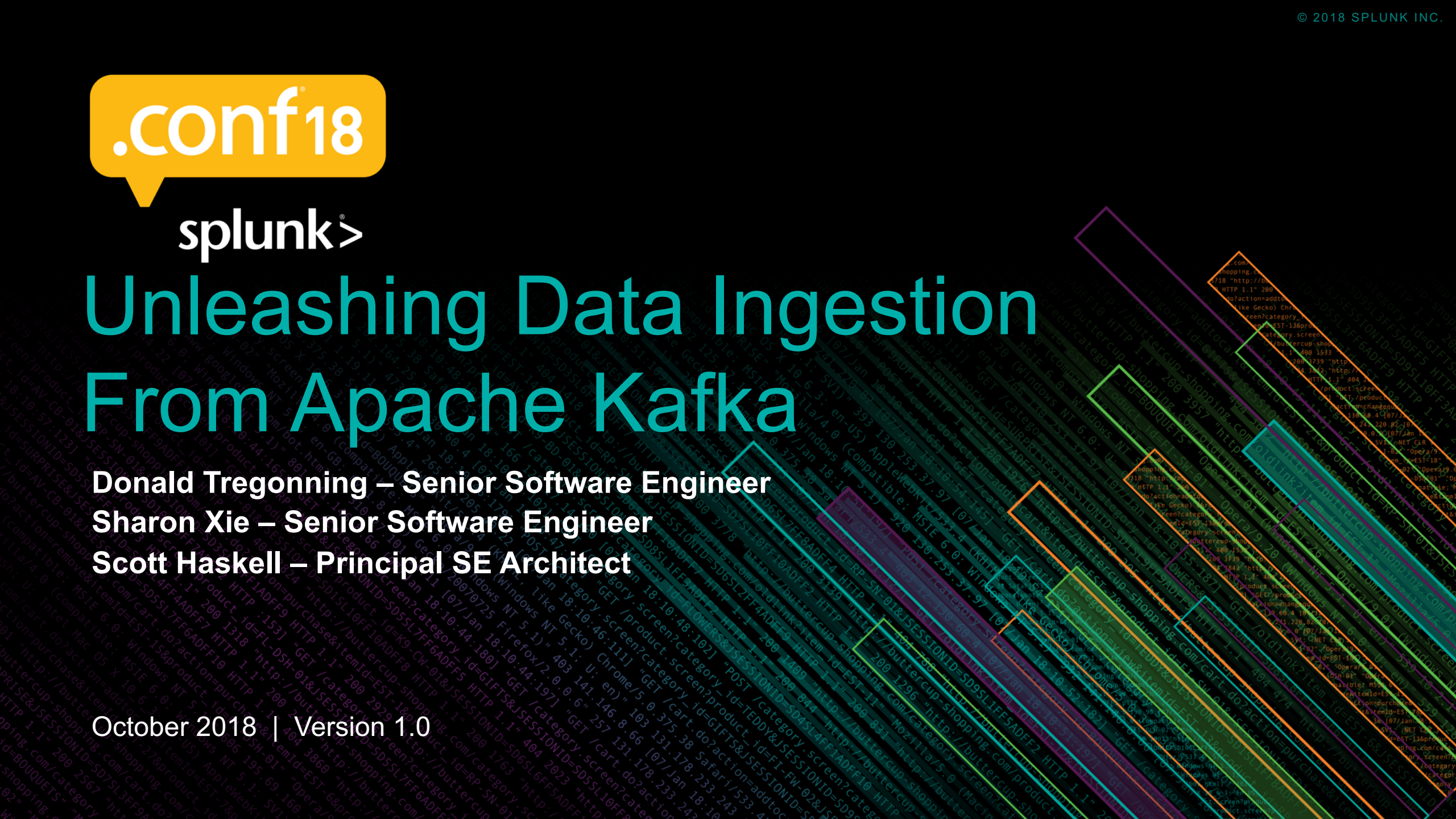


Unleashing Data Ingestion From Apache Kafka

Donald Tregonning – Senior Software Engineer
Sharon Xie – Senior Software Engineer
Scott Haskell – Principal SE Architect

October 2018 | Version 1.0

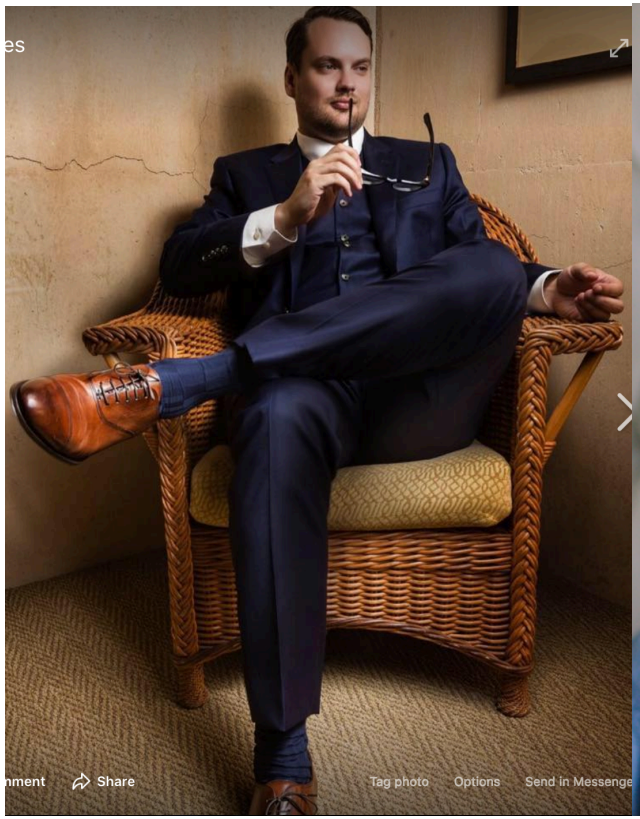
Forward-Looking Statements

During the course of this presentation, we may make forward-looking statements regarding future events or the expected performance of the company. We caution you that such statements reflect our current expectations and estimates based on factors currently known to us and that actual events or results could differ materially. For important factors that may cause actual results to differ from those contained in our forward-looking statements, please review our filings with the SEC.

The forward-looking statements made in this presentation are being made as of the time and date of its live presentation. If reviewed after its live presentation, this presentation may not contain current or accurate information. We do not assume any obligation to update any forward-looking statements we may make. In addition, any information about our roadmap outlines our general product direction and is subject to change at any time without notice. It is for informational purposes only and shall not be incorporated into any contract or other commitment. Splunk undertakes no obligation either to develop the features or functionality described or to include any such feature or functionality in a future release.

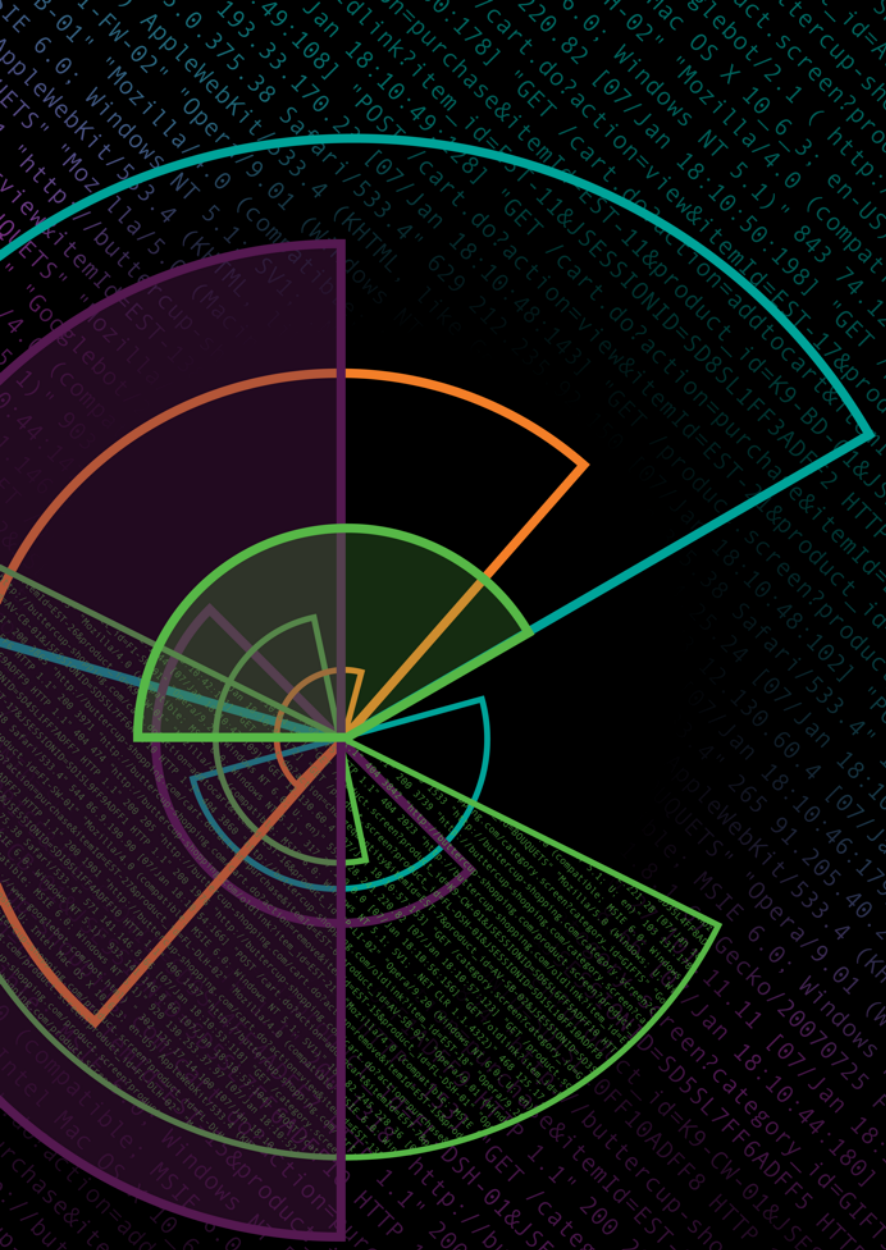
splunk> .conf18

About Us



Agenda

1. Introduction to Kafka
2. Kafka Fundamentals
3. Introduction to Splunk Connect for Kafka
4. Installation, Configuration and Deployment Demo
5. Configuration Deep Dive
6. What's New in version 1.1!

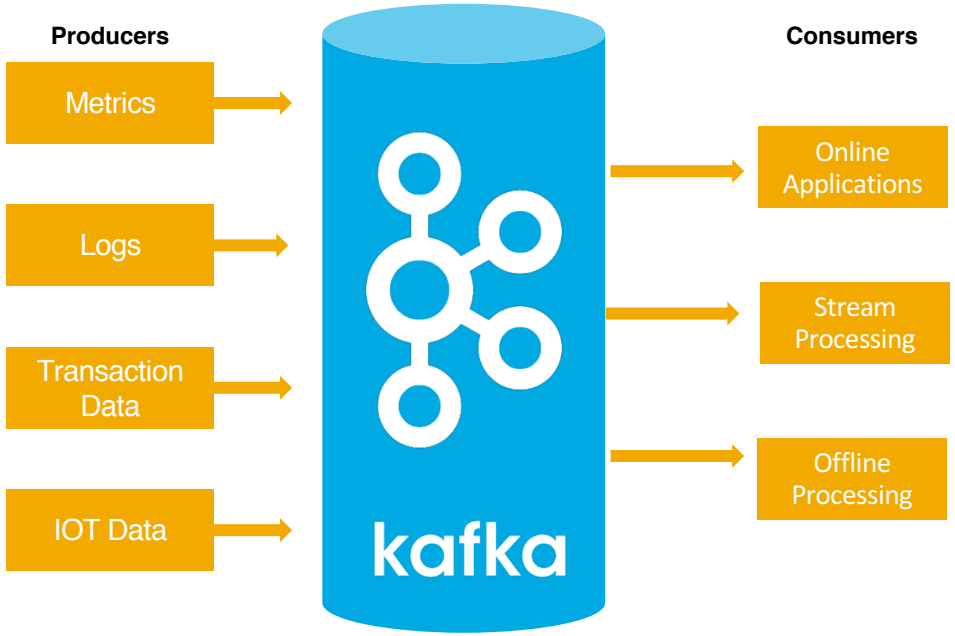


Introduction To Kafka

What is Kafka and how is it used?

Introduction to Kafka

- Apache Kafka is an open source distributed streaming platform, enabling publish/subscribe to streams of records
- Kafka is increasingly used as a foundation in data pipelines, with the largest Kafka deployments are handling over 1 trillion messages per day



- Kafka has been adopted by*:
  - 1/3 of Fortune 500
  - 7 of top 10 global banks
  - 8 of top 10 insurance companies
  - 9 of top 10 U.S. telecom companies

*Source: https://kafka.apache.org/powered-by

Companies using Kafka

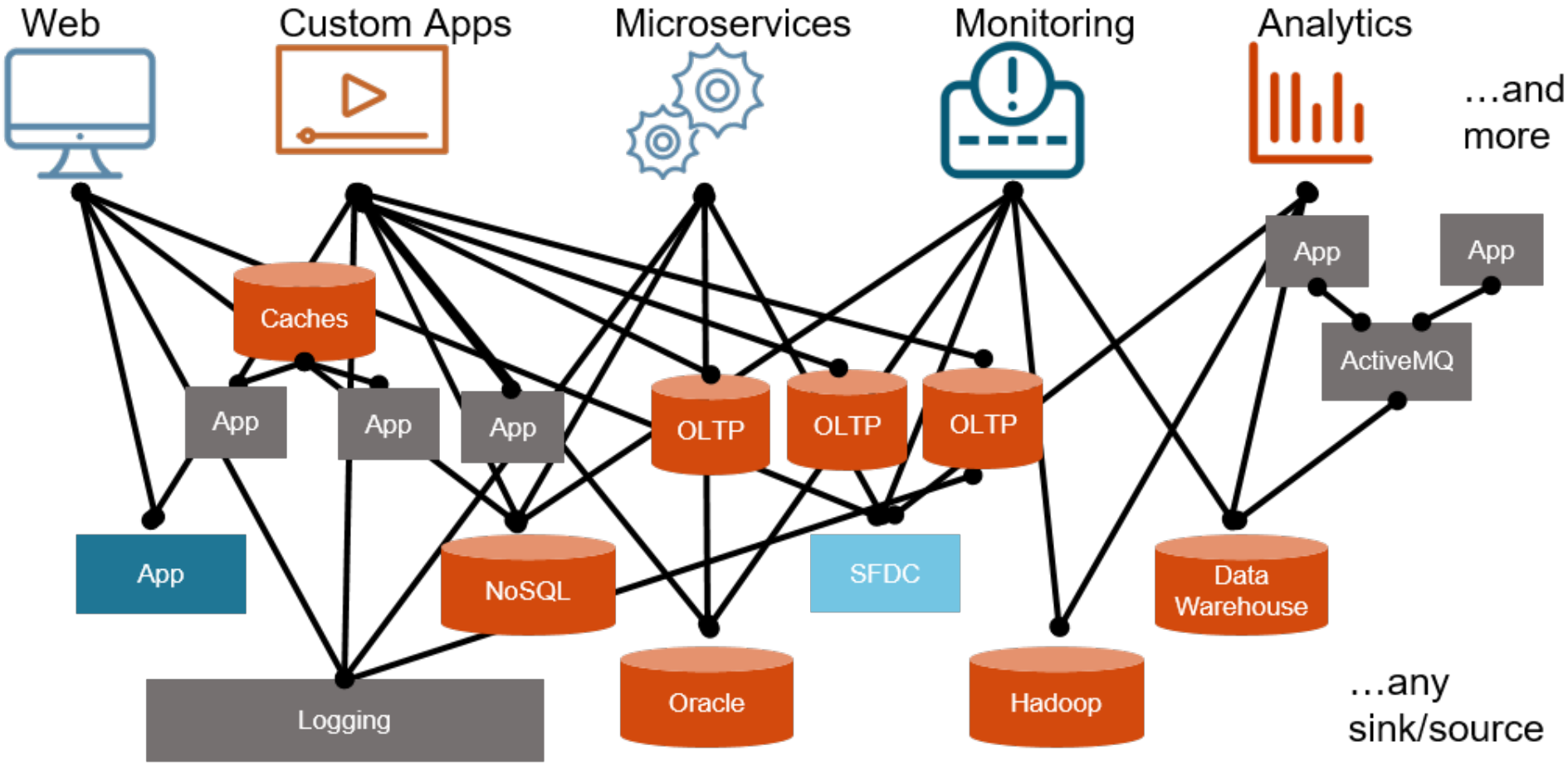
LinkedIn, Square, airbnb, Twitter, NETFLIX, The New York Times, Bloomberg

Participants in Splunk Connect for Kafka Limited Availability Release program

GitHub, citibank, PayPal, salesforce, Telstra, DHA, intel, Morgan Stanley, swisscom, splunk>, .conf18

Data transportation and messaging can be a mess..

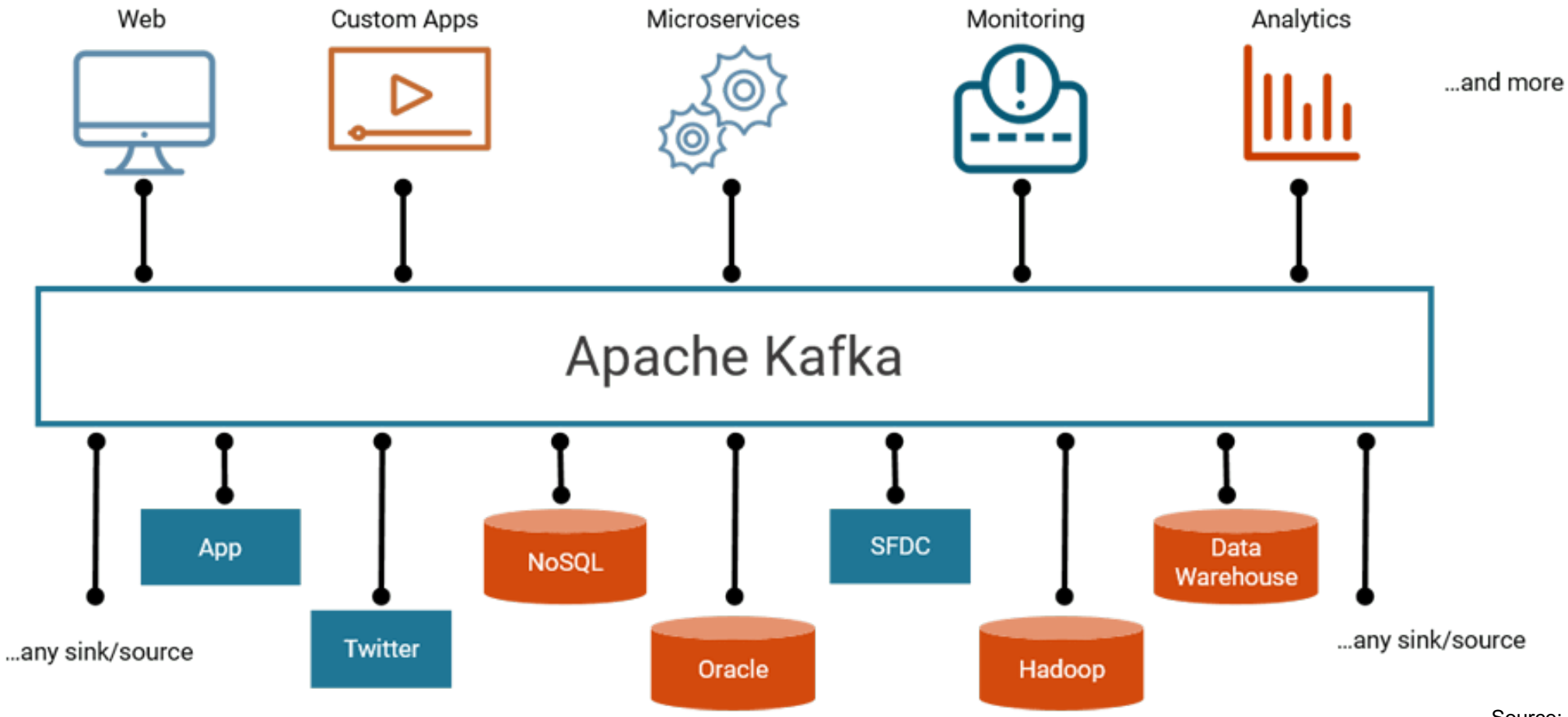
Data can be coming and going in every single direction



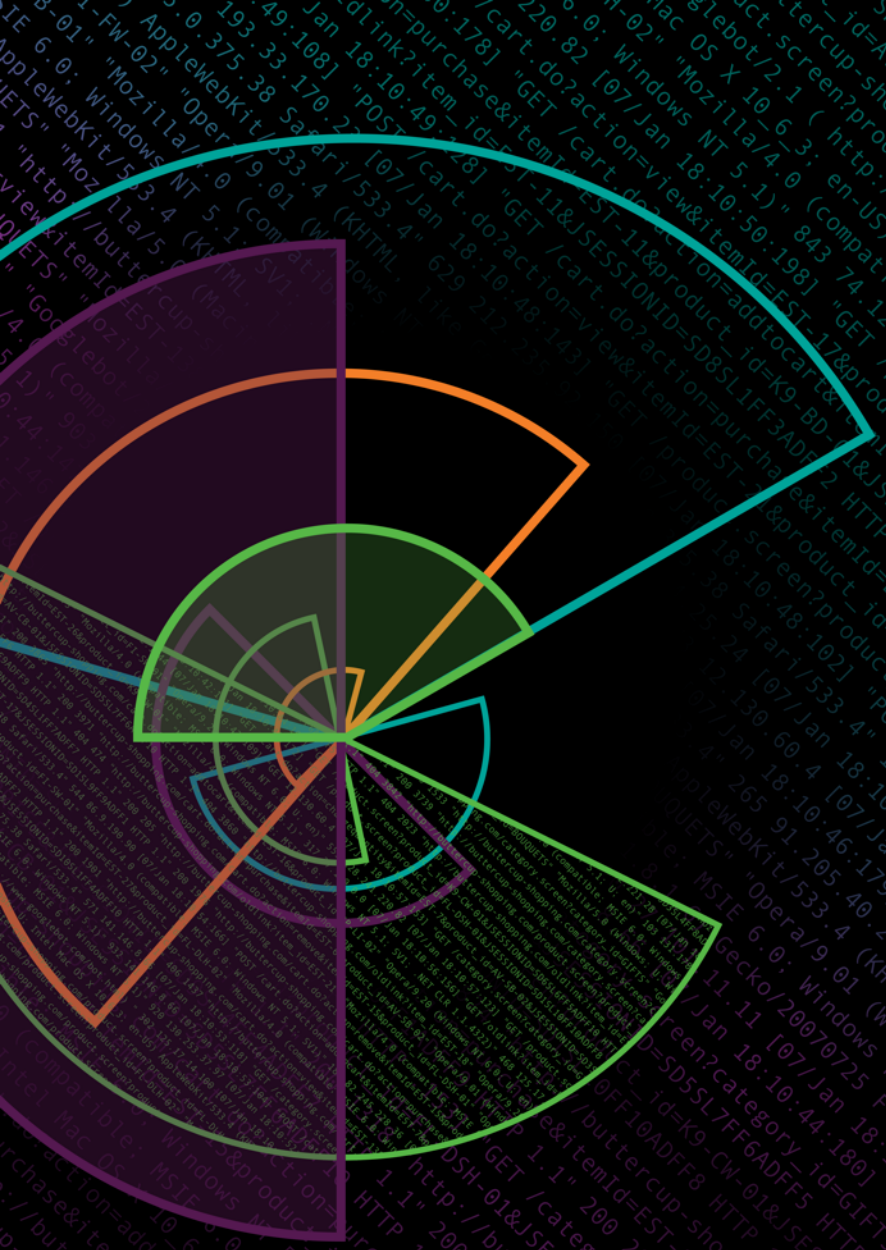
Source: Confluent.io

A common messaging bus cleans it up

With a common messaging bus.. Life gets better



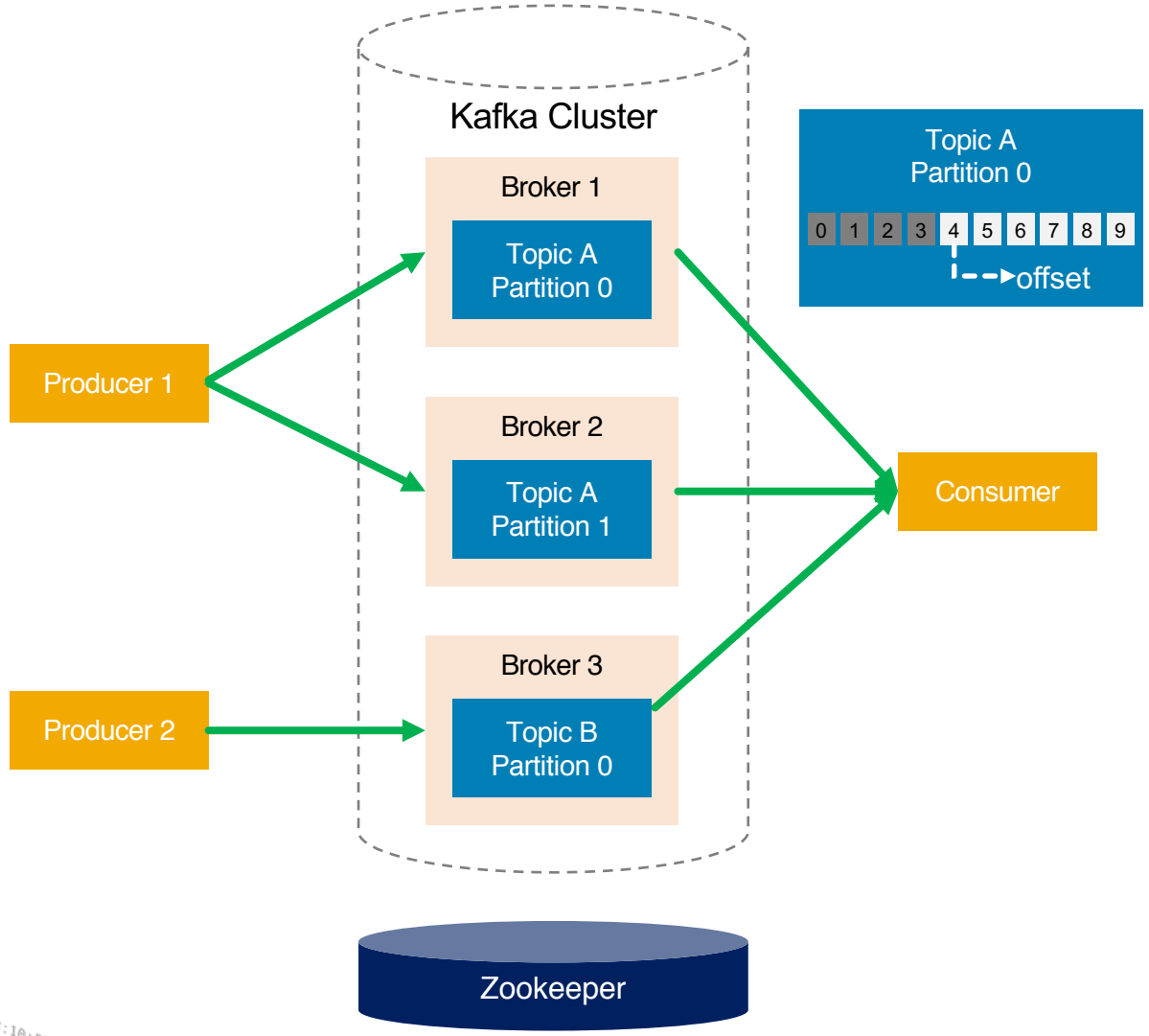
Source: Confluent.io



Kafka Fundamentals

What are topics, partitions, offsets, tasks, brokers, connectors, consumers, producers……..

Kafka Basics

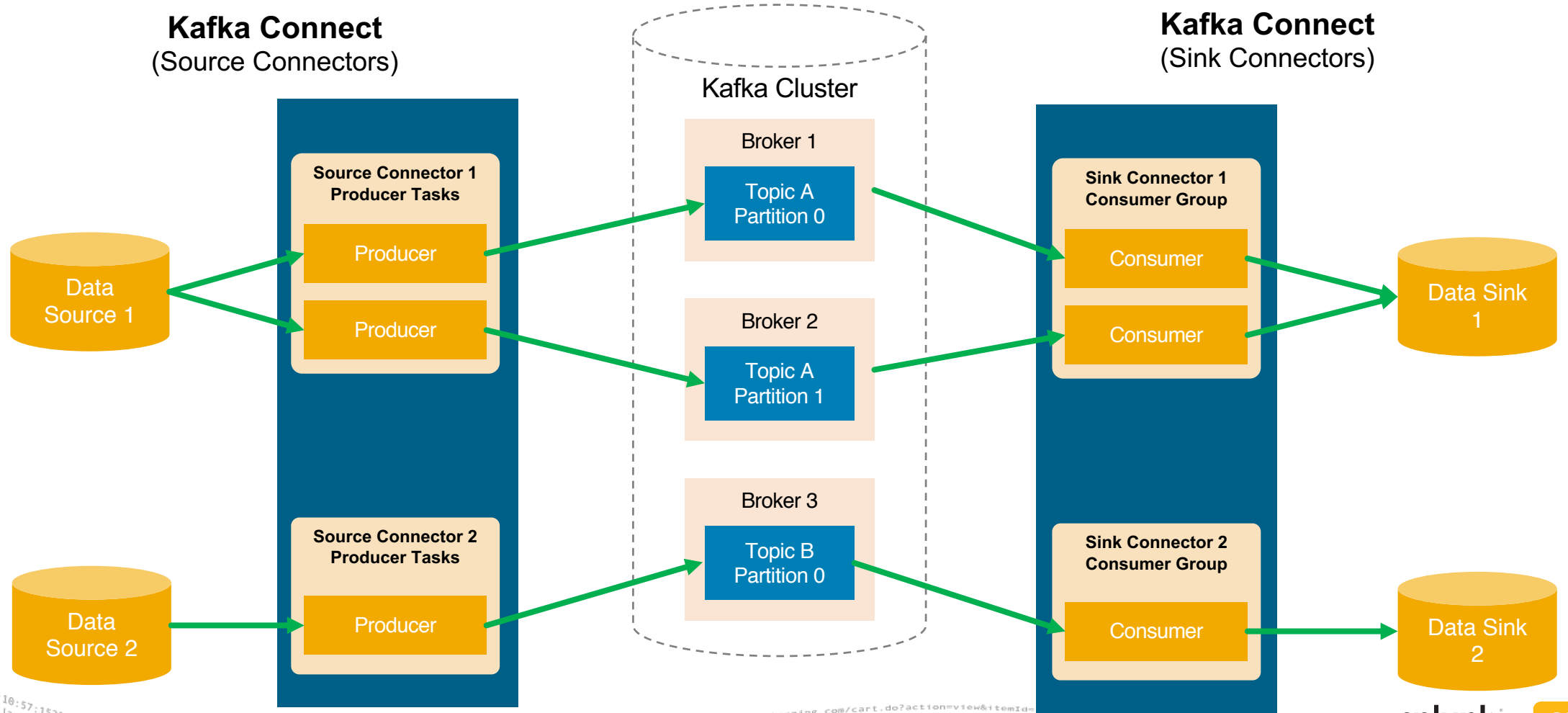


- **Producers** publish records to topics
- **Brokers** are nodes that together form the Kafka Cluster
- **Topics** are defined categories of records, similar to a table in a database or a folder in a file system
- **Partitions** are portions of a topic split among brokers to parallelize reads by consumers
- **Offsets** track progress in consuming a topic by a consumer
- **Consumers** subscribe to topics
- **Zookeeper** is used to store metadata for the brokers to enable coordination across cluster

Introduction to Kafka Connect

Kafka Connect source connectors pull data from external sources and publish to Kafka topics

Kafka Connect sink connectors subscribe to topics in Kafka and push data to external sinks



splunk> .conf18

Introduction To Splunk Connect for Kafka

The Splunk built connector:

https://github.com/splunk/kafka-connect-splunk

splunk> .conf18

Splunk Connect for Kafka

splunkbase

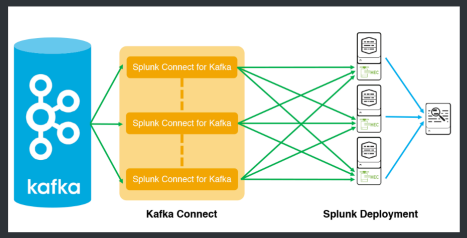
Search App by keyword, technology...

My Account | My Splunk | Support & Services

Splunk Connect for Kafka

6 ratings

Splunk Built



ADMINISTRATOR TOOLS: View App

Overview | Details

Splunk Connect for Kafka is a sink connector that allows a Splunk software administrator to subscribe to a Kafka topic and stream the data to the Splunk HTTP event collector. Built on top of the Kafka Connect library, this connector provides:
- High scalability, allowing linear scaling, limited only by the hardware supplied to the Kafka Connect environment.
- High reliability, by ensuring at-least-once delivery of data.
- Ease of data onboarding and simple configuration with Kafka Connect framework and Splunk's HTTP event collector.

Click Visit Site to download the latest release from the Splunk GitHub repository

869
Downloads

Visit Site | Rate this App

BUILT BY
Splunk Inc.

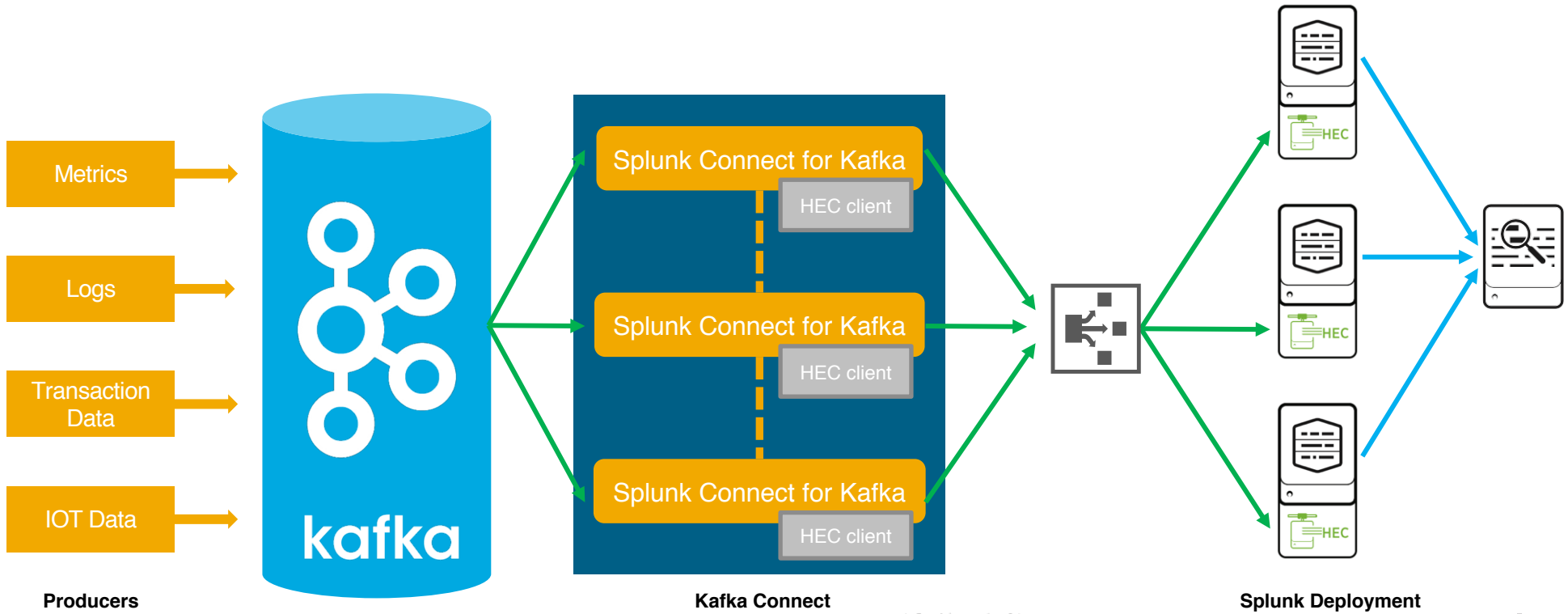
CATEGORY & CONTENTS
Categories: Utilities, DevOps
App Type: Add-on

splunk> .conf18

Splunk Connect for Kafka Overview

Splunk Connect for Kafka is a sink connector built on top of the Kafka Connect library:

- High scalability - allowing linear scaling, limited only by the hardware supplied to the Kafka Connect environment
- High reliability - ensuring at-least-once delivery of data
- High availability - deployable across a cluster allowing for outage without interruption
- Ease of data onboarding and simple configuration with Kafka Connect framework and Splunk's HTTP event collector



Why use the Splunk Connect for Kafka?

Apart from it being amazing

Officially supported Kafka to Splunk Integration

Push instead of pull using HTTP Event Collector (HEC)

Connector utilizes Kafka Connect framework allowing integration to scale

Useable across a multitude of Use Cases

splunk> .conf18

Installation, Configuration and Deployment Demo

Recorded Demo

splunk> .conf18

Set Up Kafka

1. Install Java Runtime Environment
2. Download and Install Kafka
3. Start Zookeeper
4. Start Kafka Broker
5. Create and Publish to a Topic
6. Set up HEC token
7. Install Add-Ons as Needed

Set up the Connector

1. Download and Install JAR File
2. Copy File to Connectors Directory
3. Modify Kafka Connect Properties
4. Start Kafka Connect
5. Create/Deploy the Splunk Sink Connector
6. Search for Data in Splunk

Configuration Deep Dive

Raw vs. Event HEC endpoint

- Raw
  - One Kafka record can be split into multiple Splunk events using line breaker
  - Ability to overwrite metadata (index, sourcetype, etc)

- Event
  - One Kafka record is one Splunk event
  - Ability to overwrite metadata (index, sourcetype, etc) and add custom fields.
  - Ability to send Kafka system info (eg: topic, partition, offset)

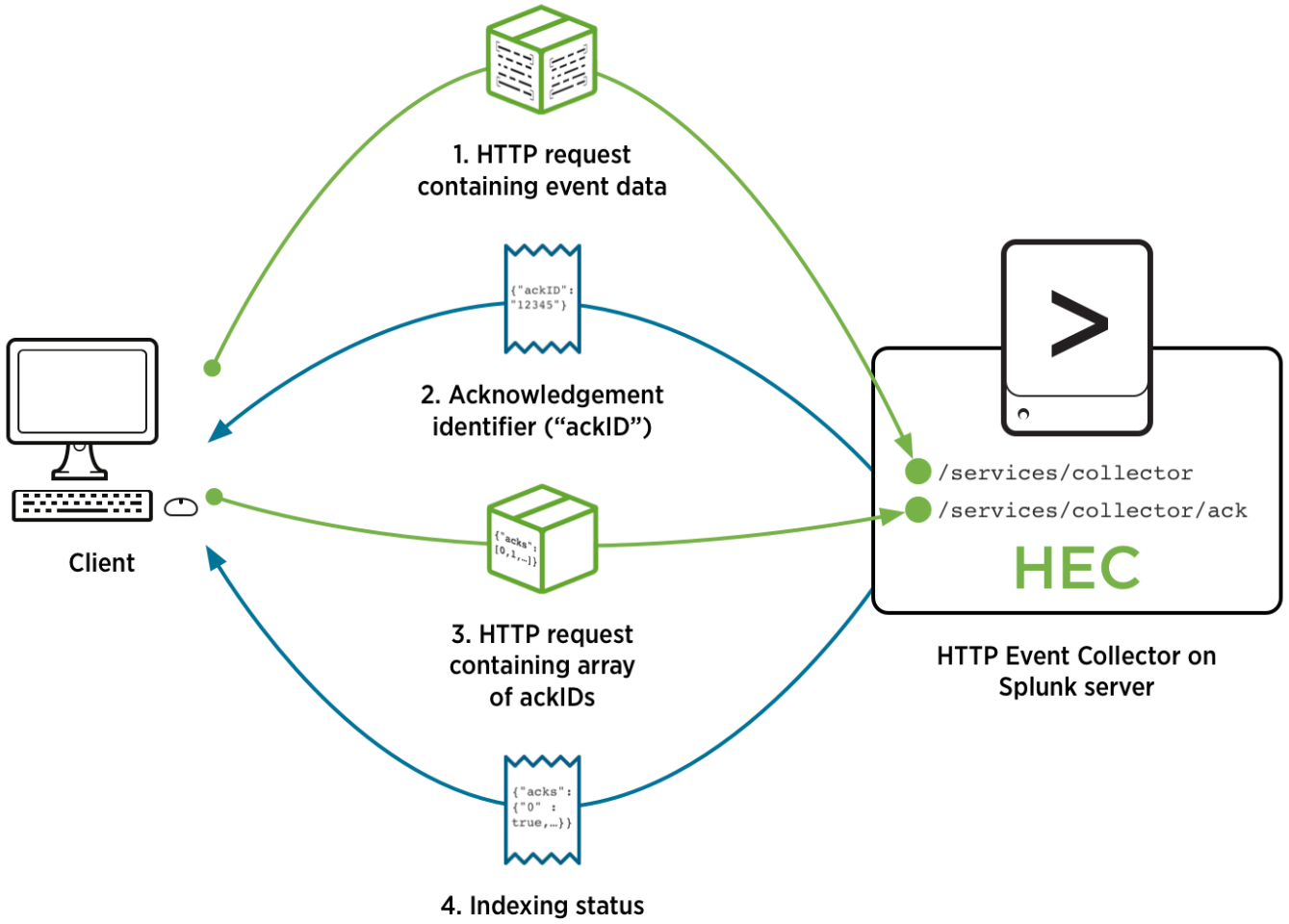
```
curl -k "https://localhost:8088/services/collector/raw?
index=main&source=datasource&sourcetype=txt&
host=localhost&timestamp=1426279439
-d '"hello": "world"'
```

```
curl -k "https://localhost:8088/services/collector/event" -d
'{
   "time": 1426279439,
   "host": "localhost",
   "source": "datasource",
   "sourcetype": "txt",
   "index": "main",
   "event": { "hello": "world" }
}'
```

splunk> .conf18

Indexer Acknowledgement

Enables “at least once delivery semantics”



Configurations (Basic)

RAW without acknowledgements, single topic

```
curl <KAFKA_CONNECT_REST_ENDPOINT>:8083/connectors -X POST -H "Content-Type: application/json" -d'{
  "name": "splunk-prod-financial",
    "config": {
      "connector.class": "com.splunk.kafka.connect.SplunkSinkConnector",
      "tasks.max": "10",
      "topics": "t1",
      "splunk.hec.uri": "https://idx1:8088,https://idx2:8088,https://idx3:8088",
      "splunk.hec.token": "1B901D2B-576D-40CD-AF1E-98141B499534",
      "splunk.hec.ack.enabled" : "false",
      "splunk.hec.raw" : "true",
      "splunk.hec.raw.line.breaker" : "#####"
    }
}'
```

Event without acknowledgements, multiple topics

```
curl <KAFKA_CONNECT_REST_ENDPOINT>:8083/connectors -X POST -H "Content-Type: application/json" -d'{
  "name": "splunk-prod-financial",
    "config": {
      "connector.class": "com.splunk.kafka.connect.SplunkSinkConnector",
      "tasks.max": "10",
      "topics": "t1,t2,t3,t4,t5,t6,t7,t8,t9,t10",
      "splunk.hec.uri": "https://idx1:8088",
      "splunk.hec.token": "1B901D2B-576D-40CD-AF1E-98141B499534",
      "splunk.hec.ack.enabled" : "false",
      "splunk.hec.raw" : "false",
      "splunk.hec.json.event.enrichment" : "org=fin,bu=south-east-us",
      "splunk.hec.track.data" : "true"

    }
}'
```

splunk> .conf18

Configurations (Acknowledgements)

RAW with additional acknowledgement configuration

```
curl <KAFKA_CONNECT_HOST>:8083/connectors -X POST -H "Content-Type: application/json" -d'{
  "name": "splunk-prod-financial",
    "config": {
      "connector.class": "com.splunk.kafka.connect.SplunkSinkConnector",
      "tasks.max": "10",
      "topics": "t1",
      "splunk.hec.uri": "https://idx1:8088",
      "splunk.hec.token": "1B901D2B-576D-40CD-AF1E-98141B499534",
      "splunk.hec.ack.enabled : "true",
      "splunk.hec.ack.poll.interval" : "20",
      "splunk.hec.ack.poll.threads" : "2",
      "splunk.hec.event.timeout" : "120",
      "splunk.hec.raw" : "true",
      "splunk.hec.raw.line.breaker" : "#####"
    }
}'
```

splunk> .conf18

Configurations (Load Balancing)

Load balancing with list of indexers

```
curl <KAFKA_CONNECT_HOST>:8083/connectors -X POST -H "Content-Type: application/json" -d'{
  "name": "splunk-prod-financial",
    "config": {
      "connector.class": "com.splunk.kafka.connect.SplunkSinkConnector",
      "tasks.max": "10",
      "topics": "t1",
      "splunk.hec.uri": "https://idx1:8088,https://idx2:8088,https://idx3:8088",
      "splunk.hec.token": "1B901D2B-576D-40CD-AF1E-98141B499534",
      "splunk.hec.ack.enabled : "true",
      "splunk.hec.raw" : "true",
      "splunk.hec.raw.line.breaker" : "#####"
    }
}'
```

Load balancing with a load balancer

```
curl <KAFKA_CONNECT_HOST>:8083/connectors -X POST -H "Content-Type: application/json" -d'{
  "name": "splunk-prod-financial",
    "config": {
      "connector.class": "com.splunk.kafka.connect.SplunkSinkConnector",
      "tasks.max": "10",
      "topics": "t1",
      "splunk.hec.uri": "https://elb-kafka:8088",
      "splunk.hec.token": "1B901D2B-576D-40CD-AF1E-98141B499534",
      "splunk.hec.ack.enabled : "true",
      "splunk.hec.raw" : "true",
      "splunk.hec.raw.line.breaker" : "#####”
      "splunk.hec.total.channels": "4"
    }
}'
```

splunk> .conf18

Configurations (Metrics)

collectd and RAW, for use with collectd_http pre-trained sourcetype in Splunk

```
curl <hostname>:8083/connectors -X POST -H "Content-Type: application/json" -d'{
  "name": "splunk-prod-financial",
    "config": {
      "connector.class": "com.splunk.kafka.connect.SplunkSinkConnector",
      "tasks.max": "10",
      "topics": "t1,t2,t3,t4,t5,t6,t7,t8,t9,t10",
      "splunk.sourcetypes": "collectd_http",
      "splunk.hec.uri": "https://idx1:8088,https://idx2:8088,https://idx3:8088",
      "splunk.hec.token": "1B901D2B-576D-40CD-AF1E-98141B499534",
      "splunk.hec.ack.enabled : "true",
      "splunk.hec.ack.poll.interval" : "20",
      "splunk.hec.ack.poll.threads" : "2",
      "splunk.hec.event.timeout" : "120",
      "splunk.hec.raw" : "true",
      "splunk.hec.raw.line.breaker" : "#####"
    }
}'
```

Configurations (Topic based Index Routing)

Topic-to-Index routing defined in connector configuration

```
curl <KAFKA_CONNECT_REST_ENDPOINT>:8083/connectors -X POST -H "Content-Type: application/json" -d'{
  "name": "splunk-prod-financial",
    "config": {
      "connector.class": "com.splunk.kafka.connect.SplunkSinkConnector",
      "tasks.max": "10",
      "topics": "test-1,test-2,test-3”,
      "splunk.indexes": "kafka-1,kafka-2,kafka-3"
      "splunk.hec.uri": "https://idx1:8088,https://idx2:8088,https://idx3:8088",
      "splunk.hec.token": "1B901D2B-576D-40CD-AF1E-98141B499534",
      "splunk.hec.ack.enabled : "false",
      "splunk.hec.raw" : "true",
      "splunk.hec.raw.line.breaker" : "#####"
    }
}'
```

splunk> .conf18

Configurations (Event based Index Routing)

Splunk configuration

props.conf on indexers

```
[kafka:events]
TRANSFORMS-index_routing = route_data_to_index_by_field_owner_id
```

transforms.conf on indexers

```
[route_data_to_index_by_field_owner_id]
REGEX = "(\w+)":"123456789012"
DEST_KEY = _MetaData:Index
FORMAT = prod
```

Example event

```
{
    "owner": "123456789012",
    "logGroup": "CloudTrail",
    "logStream": "123456789012_CloudTrail_us-east-1",
    "subscriptionFilters": ["Destination"],
    "messageType": "DATA_MESSAGE",
    "logEvents": {
        "id": "31953106606966983378809025079804211143289615424298221570",
        "timestamp": 1432826855000,
        "message": {
            "eventVersion": "1.03",
            "userIdentity": {
                "type": "Root"
            }
        }
    }
}
```

splunk> .conf18

Configuration (Event based Index Routing using Headers) – New in 1.1!!

```
curl <hostname>:8083/connectors -X POST -H "Content-Type: application/json" -d'{
  "name": "splunk-prod-financial",
    "config": {
      "connector.class": "com.splunk.kafka.connect.SplunkSinkConnector",
      "tasks.max": "10",
      "topics": "t1,t2,t3,t4,t5,t6,t7,t8,t9,t10",
      "splunk.sourcetypes": "collectd_http",
      "splunk.hec.uri": "https://idx1:8088,https://idx2:8088,https://idx3:8088",
      "splunk.hec.token": "1B901D2B-576D-40CD-AF1E-98141B499534",
      "splunk.hec.ack.enabled : "true",
      "splunk.hec.ack.poll.interval" : "20",
      "splunk.hec.ack.poll.threads" : "2",
      "splunk.hec.event.timeout" : "120",
      "splunk.hec.raw" : "false",
      "splunk.header.support": "true",
      "splunk.header.index" : ”destination_storage",
      "splunk.header.source" : ”Financial_Application",
      "splunk.header.sourcetype" : ”ledger_format",
      "splunk.header.host": ”finance.company.host"
      }
}'
```

splunk> .conf18

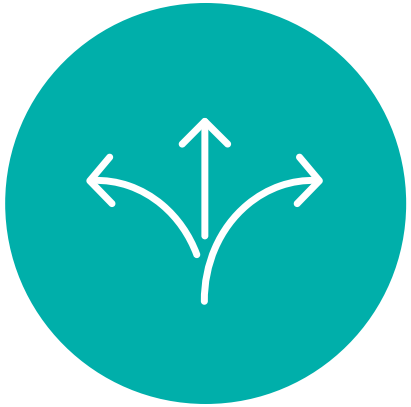
What's New in Version 1.1

Released September 2018

splunk> .conf18

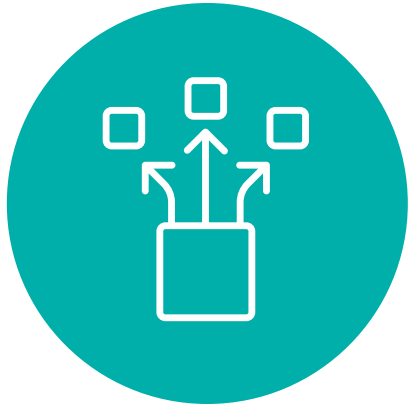
Splunk Connect for Kafka Version 1.1

Structed Data (Avro, JSON)



Per-Event Routing / Metadata Override with Kafka Headers

HEC Event Format



Sticky Session Improvements

Custom Java Keystore Location

Key Learning Takeaways

New Push-based way to get data from Kafka into Splunk

- Kafka is increasingly important for customers building large data pipelines
- Splunk's new connector uses Kafka Connect framework for horizontally scaling consumption of records from Kafka and pushing them to Splunk through HEC

Connector provides flexibility to scale usage

- HEC is used, and customers can load balance across Splunk indexers specifying a list of indexers or a load balancer
- Indexer acknowledgement guarantees at-least-once delivery
- Index routing allows separation of data from topics to different indexes
- Metric store support using collectd and raw mode (collectd_http sourcetype)

Splunk Add-on for Kafka modular input is deprecated

- Modular input for consuming topics is now deprecated, but still enabled
- Kafka environment monitoring using JMX will remain as is

Documentation:
http://docs.splunk.com/Documentation/KafkaConnect

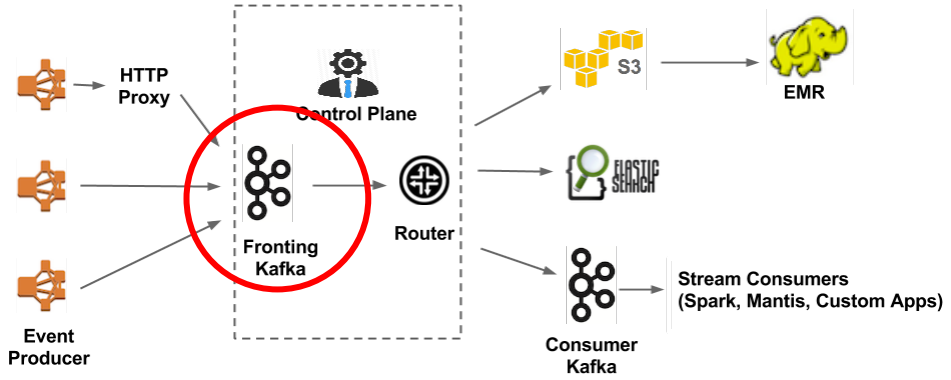
splunk> .conf18

Don't forget to rate this session
in the .conf18 mobile app

Reference Slides

Kafka in Practice

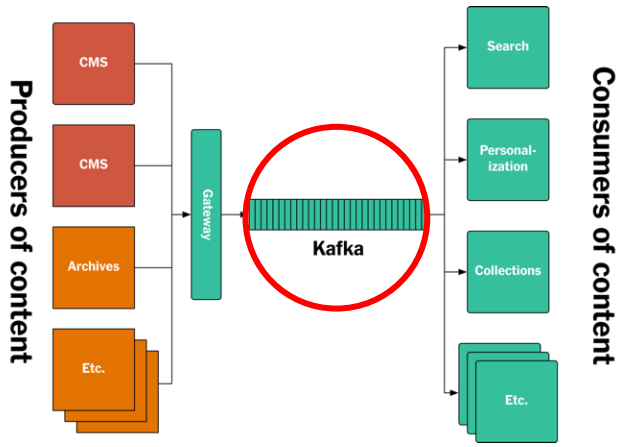
NETFLIX



- 36 Kafka clusters consisting of 4,000+ broker instances
- > 700 billion messages / day on average

Source: Netflix, 2016

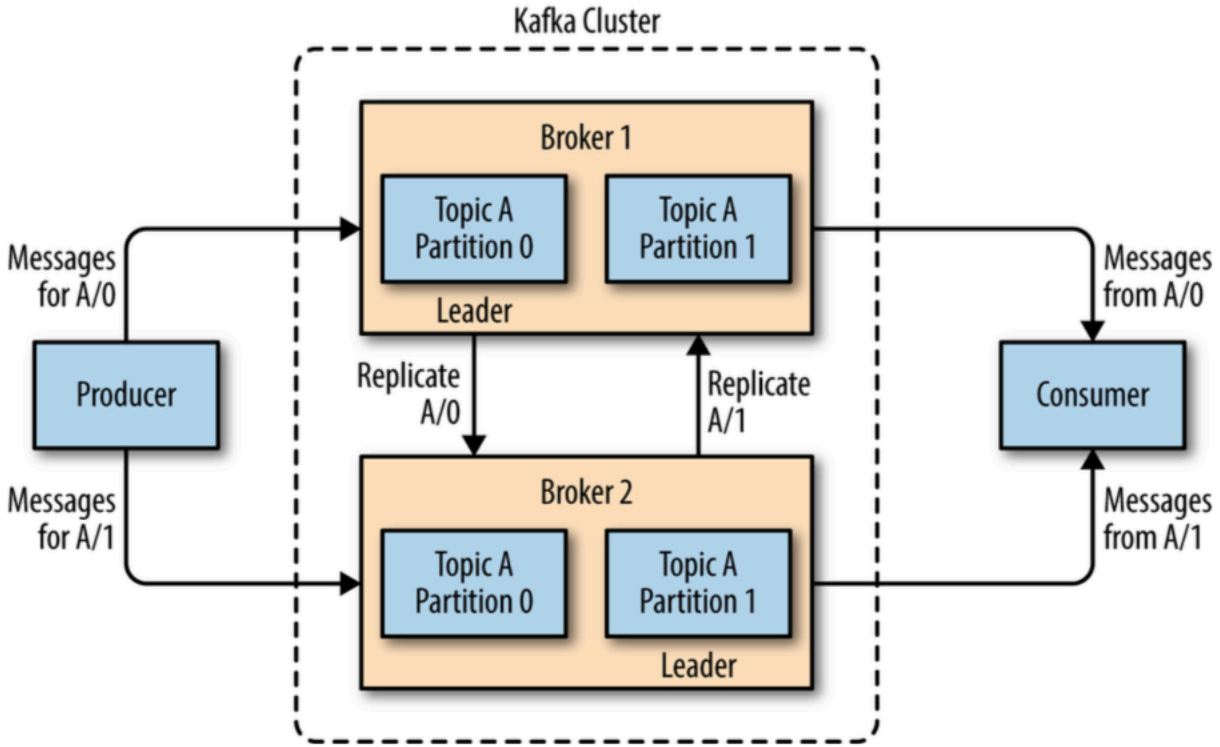
The New York Times



- Need to retain all events forever to recreate a data store from scratch
- Log consumption must be be ordered – if events with causal relationships are processed out of order, the result will be wrong

Source: New York Times, 2017

Kafka Basics (Cont..)



Add Detail Slides For Reference Here

- Reference specs

Installation, Configuration and Deployment Demo

splunk> .conf18

Set Up Kafka

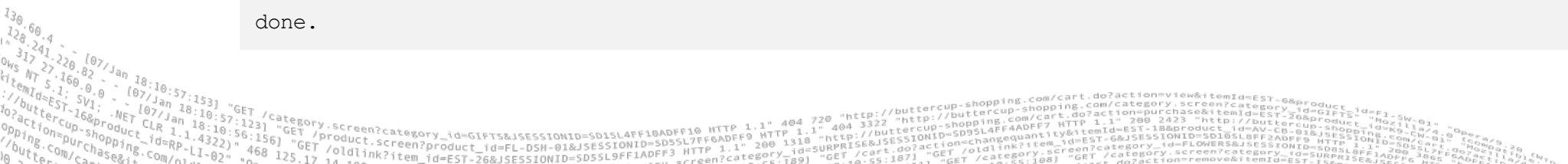
1. Install Java Runtime Environment
2. Download and Install Kafka
3. Start Zookeeper
4. Start Kafka Broker
5. Create and Publish to a Topic
6. Set up HEC token
7. Install Add-Ons as Needed

Set up the Connector

1. Download and Install JAR File
2. Copy File to Connectors Directory
3. Modify Kafka Connect Properties
4. Start Kafka Connect
5. Create/Deploy the Splunk Sink Connector
6. Search for Data in Splunk

Install Java Runtime Environment

```
kafka@ip-172-31-21-188:/opt$ apt-get update
Hit:1 http://us-east-2.ec2.archive.ubuntu.com/ubuntu xenial InRelease
Get:2 http://us-east-2.ec2.archive.ubuntu.com/ubuntu xenial-updates InRelease [102 kB]
Get:3 http://us-east-2.ec2.archive.ubuntu.com/ubuntu xenial-backports InRelease [102 kB]
Get:4 http://us-east-2.ec2.archive.ubuntu.com/ubuntu xenial/main Sources [868 kB]
…
Get:38 http://security.ubuntu.com/ubuntu xenial-security/multiverse Translation-en [1,408 B]
Fetched 25.0 MB in 6s (3,767 kB/s)
Reading package lists... Done

kafka@ip-172-31-21-188:/opt$ apt-get install default-jre
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following additional packages will be installed:
  ca-certificates-java default-jre-headless fontconfig fontconfig-config fonts-dejavu-core fonts-dejavu-extra
hicolor-icon-theme java-common libasound2 libasound2-data libasyncns0
  libatk1.0-0 libatk1.0-data libavahi-client3 libavahi-common-data libavahi-common3 libcairo2 libcups2 libdatrie1
…
Setting up openjdk-8-jre-headless:amd64 (8u151-b12-0ubuntu0.16.04.2) ...
update-alternatives: using /usr/lib/jvm/java-8-openjdk-amd64/jre/bin/rmid to provide /usr/bin/rmid (rmid) in auto
mode
update-alternatives: using /usr/lib/jvm/java-8-openjdk-amd64/jre/bin/keytool to provide /usr/bin/keytool
…
Processing triggers for ca-certificates (20170717~16.04.1) ...
Updating certificates in /etc/ssl/certs...
0 added, 0 removed; done.
Running hooks in /etc/ca-certificates/update.d...

done.
```



splunk> .conf18

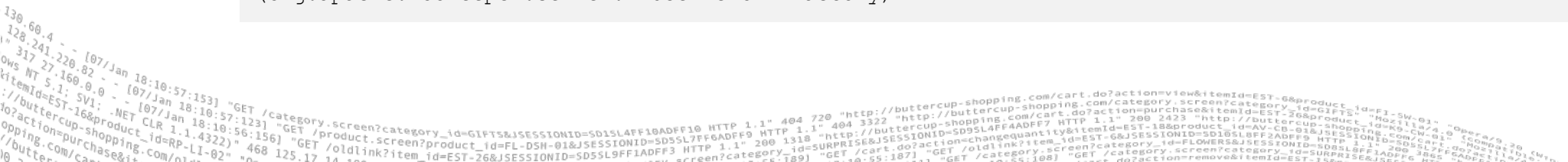
Download and Install Kafka

```
kafka@ip-172-31-21-188:/opt$ wget http://apache.claz.org/kafka/1.0.1/kafka_2.11-1.0.1.tgz
--2018-03-16 05:00:25--  http://apache.claz.org/kafka/1.0.1/kafka_2.11-1.0.1.tgz
Resolving apache.claz.org (apache.claz.org)... 74.63.227.45
Connecting to apache.claz.org (apache.claz.org)|74.63.227.45|:80... connected.
HTTP request sent, awaiting response... 200 OK
Length: 49766096 (47M) [application/x-gzip]
Saving to: ‘kafka_2.11-1.0.1.tgz’

kafka_2.11-1.0.1.tgz
100%[=======================================================================================>]  47.46M
10.8MB/s    in 4.6s

2018-03-16 05:00:30 (10.3 MB/s) - ‘kafka_2.11-1.0.1.tgz’ saved [49766096/49766096]

kafka@ip-172-31-21-188:/opt$ tar xvf kafka_2.11-1.0.1.tgz
kafka_2.11-1.0.1/
kafka_2.11-1.0.1/LICENSE
kafka_2.11-1.0.1/NOTICE
kafka_2.11-1.0.1/bin/
kafka_2.11-1.0.1/bin/kafka-delete-records.sh
kafka_2.11-1.0.1/bin/trogdor.sh
kafka_2.11-1.0.1/bin/kafka-preferred-replica-ele
…
kafka@ip-172-31-21-188:/opt$ mv kafka_2.11-1.0.1 kafka
```

splunk> .conf18

Start Zookeeper

```
kafka@ip-172-31-21-188:/opt/kafka$ $KAFKA_HOME/bin/zookeeper-server-start.sh
$KAFKA_HOME/config/zookeeper.properties
[2018-03-16 05:15:36,823] INFO Reading configuration from: /opt/kafka/config/zookeeper.properties
(org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2018-03-16 05:15:36,825] INFO autopurge.snapRetainCount set to 3
(org.apache.zookeeper.server.DatadirCleanupManager)
[2018-03-16 05:15:36,825] INFO autopurge.purgeInterval set to 0
(org.apache.zookeeper.server.DatadirCleanupManager)
[2018-03-16 05:15:36,825] INFO Purge task is not scheduled. (org.apache.zookeeper.server.DatadirCleanupManager)
[2018-03-16 05:15:36,825] WARN Either no config or no quorum defined in config, running  in standalone mode
(org.apache.zookeeper.server.quorum.QuorumPeerMain)
[2018-03-16 05:15:36,836] INFO Reading configuration from: /opt/kafka/config/zookeeper.properties
(org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2018-03-16 05:15:36,837] INFO Starting server (org.apache.zookeeper.server.ZooKeeperServerMain)
[2018-03-16 05:15:36,862] INFO Server environment:zookeeper.version=3.4.10-
39d3a4f269333c922ed3db283be479f9deacaa0f, built on 03/23/2017 10:13 GMT
(org.apache.zookeeper.server.ZooKeeperServer)
[2018-03-16 05:15:36,862] INFO Server environment:host.name=ip-172-31-21-188.us-east-2.compute.internal
(org.apache.zookeeper.server.ZooKeeperServer)
[2018-03-16 05:15:36,863] INFO Server environment:java.version=1.8.0_151
(org.apache.zookeeper.server.ZooKeeperServer)
…
[2018-03-16 05:15:36,863] INFO Server environment:user.dir=/opt/kafka
(org.apache.zookeeper.server.ZooKeeperServer)
[2018-03-16 05:15:36,869] INFO tickTime set to 3000 (org.apache.zookeeper.server.ZooKeeperServer)
[2018-03-16 05:15:36,869] INFO minSessionTimeout set to -1 (org.apache.zookeeper.server.ZooKeeperServer)
[2018-03-16 05:15:36,869] INFO maxSessionTimeout set to -1 (org.apache.zookeeper.server.ZooKeeperServer)
[2018-03-16 05:15:36,876] INFO binding to port 0.0.0.0/0.0.0.0:2181
(org.apache.zookeeper.server.NIOServerCnxnFactory)
```



splunk> .conf18

Start Kafka Broker

```
kafka@ip-172-31-21-188:/opt/kafka$ $KAFKA_HOME/bin/kafka-server-start.sh $KAFKA_HOME/config/server.properties
[2018-03-16 05:19:48,864] INFO KafkaConfig values:
            advertised.host.name = null
            advertised.listeners = null
            advertised.port = null
            alter.config.policy.class.name = null
            authorizer.class.name =
            auto.create.topics.enable = true
            auto.leader.rebalance.enable = true
            background.threads = 10
            broker.id = 0
            broker.id.generation.enable = true
            broker.rack = null
            compression.type = producer
            connections.max.idle.ms = 600000
…
[2018-03-16 05:19:49,559] INFO [TransactionCoordinator id=0] Starting up.
(kafka.coordinator.transaction.TransactionCoordinator)
[2018-03-16 05:19:49,560] INFO [Transaction Marker Channel Manager 0]: Starting
(kafka.coordinator.transaction.TransactionMarkerChannelManager)
[2018-03-16 05:19:49,560] INFO [TransactionCoordinator id=0] Startup complete.
(kafka.coordinator.transaction.TransactionCoordinator)
[2018-03-16 05:19:49,582] INFO Creating /brokers/ids/0 (is it secure? false) (kafka.utils.ZKCheckedEphemeral)
[2018-03-16 05:19:49,585] INFO Result of znode creation is: OK (kafka.utils.ZKCheckedEphemeral)
[2018-03-16 05:19:49,586] INFO Registered broker 0 at path /brokers/ids/0 with addresses: EndPoint(ip-172-31-21-
188.us-east-2.compute.internal,9092,ListenerName(PLAINTEXT),PLAINTEXT) (kafka.utils.ZkUtils)
[2018-03-16 05:19:49,587] WARN No meta.properties file under dir /tmp/kafka-logs/meta.properties
(kafka.server.BrokerMetadataCheckpoint)
[2018-03-16 05:19:49,596] INFO Kafka version : 1.0.1 (org.apache.kafka.common.utils.AppInfoParser)
[2018-03-16 05:19:49,596] INFO Kafka commitId : c0518aa65f25317e (org.apache.kafka.common.utils.AppInfoParser)
```

splunk> .conf18

Create a Kafka Topic

Create a new topic

```
kafka@ip-172-31-21-188:/opt/kafka$ $KAFKA_HOME/bin/kafka-topics.sh --create --zookeeper localhost:2181 --
replication-factor 1 --partitions 1 --topics cisco_asa
WARNING: Due to limitations in metric names, topics with a period ('.') or underscore ('') could collide. To
avoid issues it is best to use either, but not both.
Created topic "cisco_asa"
```

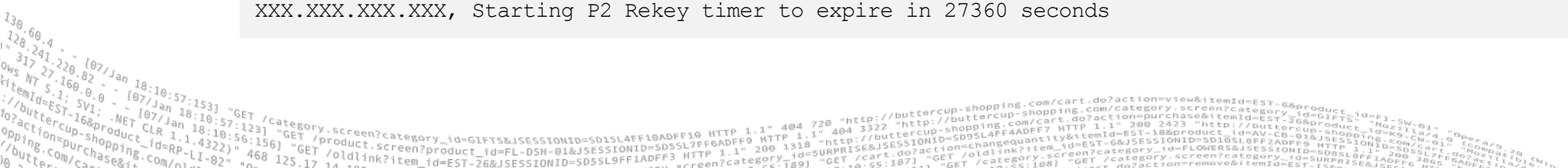
List available topics to verify topic is accessible

```
kafka@ip-172-31-21-188:/opt/kafka$ $KAFKA_HOME/bin/kafka-topics.sh --list --zookeeper localhost:2181
cisco_asa
```

splunk> .conf18

Publish Events to Topic

```
kafka@ip-172-31-21-188:/opt/kafka$ $KAFKA_HOME/bin/kafka-console-producer.sh --broker-list --topic cisco_asa
Mar 15 12:00:22 XXX.XXX.XXX.XXX %ASA-4-400013 IPS:2003 ICMP redirect from XXX.XXX.XXX.XXX to XXX.XXX.XXX.XXX on
interface dmz
Mar 15 12:00:24 XXX.XXX.XXX.XXX %ASA-6-302013: Built outbound TCP connection 9 for outside:XXX.XXX.XXX.XXX/22
(XXX.XXX.XXX.XXX/22) to inside:XXX.XXX.XXX.XXX/53496 (XXX.XXX.XXX.XXX/53496)
Mar 15 12:01:24 XXX.XXX.XXX.XXX %ASA-6-109025: Authorization denied (acl=acmetechinbound) for user 'UUUUUUUU'
from XXX.XXX.XXX.XXX/64857 to XXX.XXX.XXX.XXX/53 on interface Outside using UDP
Mar 15 12:01:24 XXX.XXX.XXX.XXX %ASA-6-302013: Built inbound TCP connection 518029 for
Outside:XXX.XXX.XXX.XXX/1123 (XXX.XXX.XXX.XXX/1123) to Inside:XXX.XXX.XXX.XXX/8443 (XXX.XXX.XXX.XXX/8443)
(UUUUUUUU)
Mar 15 12:01:24 XXX.XXX.XXX.XXX %ASA-6-302014: Teardown TCP connection 518028 for Outside:XXX.XXX.XXX.XXX/1122 to
Inside:XXX.XXX.XXX.XXX/8443 duration 0:00:03 bytes 1449 TCP FINs (UUUUUUUU)
Mar 15 12:01:24 XXX.XXX.XXX.XXX %ASA-6-109025: Authorization denied (acl=acmetechoutbound) for user 'UUUUUUUU'
from XXX.XXX.XXX.XXX/1118 to XXX.XXX.XXX.XXX/80 on interface Outside using TCP
Mar 15 12:01:21 XXX.XXX.XXX.XXX %ASA-6-302014: Teardown TCP connection 518026 for Outside:XXX.XXX.XXX.XXX/1120 to
Inside:XXX.XXX.XXX.XXX/8443 duration 0:00:03 bytes 932 TCP FINs (UUUUUUUU)
Mar 15 12:00:37 XXX.XXX.XXX.XXX %ASA-6-302016: Teardown UDP connection 517966 for Inside:XXX.XXX.XXX.XXX/30357 to
NP Identity Ifc:XXX.XXX.XXX.XXX/161 duration 0:02:01 bytes 1055
Mar 15 12:00:05 XXX.XXX.XXX.XXX %ASA-3-713119: Group = Acme_techoutbound, Username = UUUUUUUU, IP =
XXX.XXX.XXX.XXX, PHASE 1 COMPLETED
Mar 15 12:00:05 XXX.XXX.XXX.XXX %ASA-5-713075: Group = Acme_techoutbound, Username = UUUUUUUU, IP =
XXX.XXX.XXX.XXX, Overriding Initiator's IPSec rekeying duration from 2147483 to 28800 seconds
Mar 15 12:00:05 XXX.XXX.XXX.XXX %ASA-6-602303: IPSEC: An outbound remote access SA (SPI= 0x99DAB075) between
XXX.XXX.XXX.XXX and XXX.XXX.XXX.XXX (user= UUUUUUUU) has been created.
Mar 15 12:00:05 XXX.XXX.XXX.XXX %ASA-5-713049: Group = Acme_techoutbound, Username = UUUUUUUU, IP =
XXX.XXX.XXX.XXX, Security negotiation complete for User (UUUUUUUU) Responder, Inbound SPI = 0xae9ee6c6, Outbound
SPI = 0x99dab075
Mar 15 12:00:05 XXX.XXX.XXX.XXX %ASA-6-713905: Group = Acme_techoutbound, Username = UUUUUUUU, IP =
XXX.XXX.XXX.XXX, Starting P2 Rekey timer to expire in 27360 seconds
```



splunk> .conf18

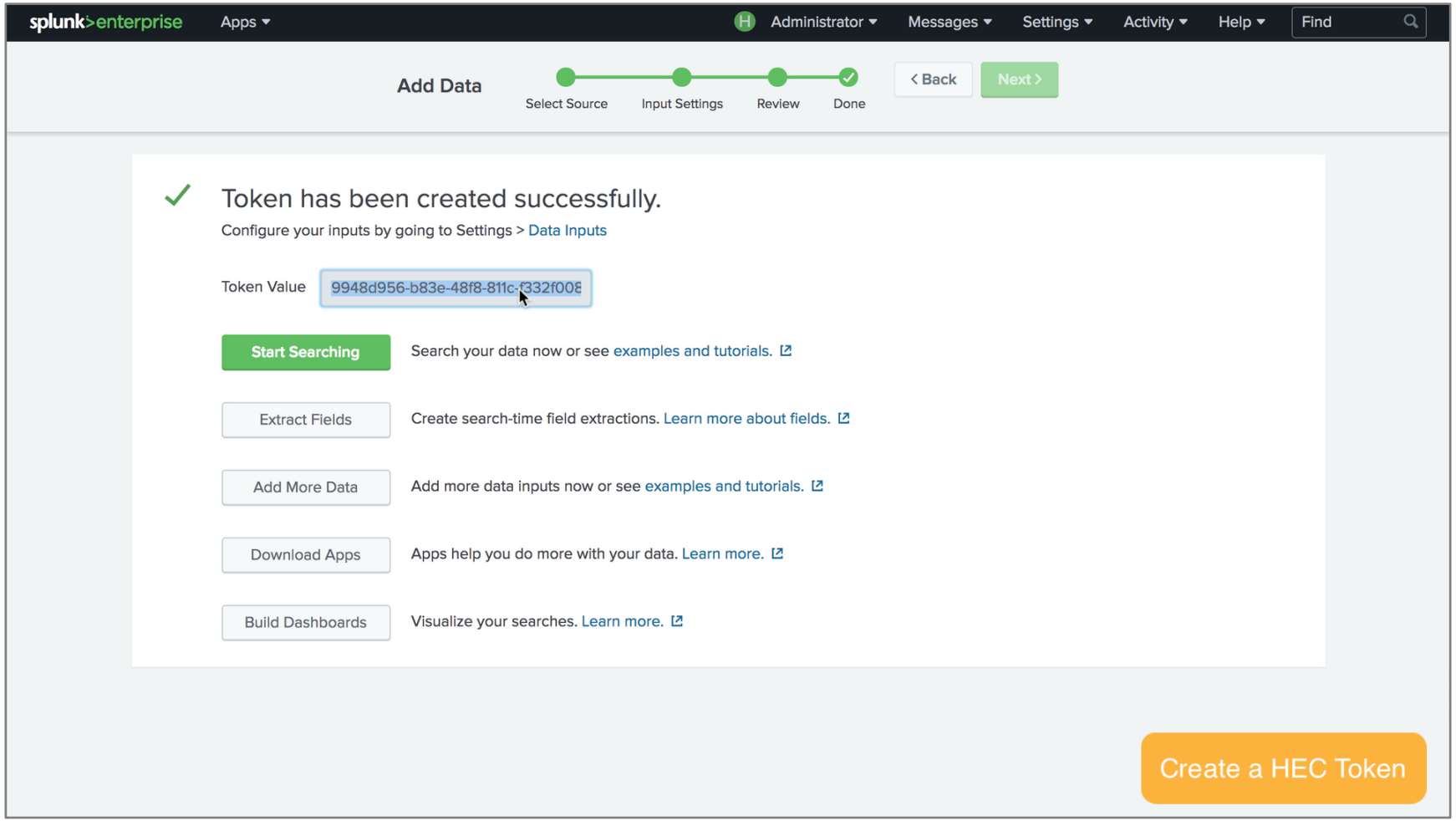
Configure Splunk Environment

1. Set up HEC token
2. Install Add-Ons as Needed

HTTP Event Collector

- Create a new token, and replicate that on all indexers or heavy forwarders, whichever will be receiving data from Kafka Connect
- For guaranteed at-least-once delivery of data, you can enable indexer acknowledgements for the HEC token. This will also require setting `splunk.hec.ack.enabled` to true when configuring the connector

HTTP Event Collector (Cont..)



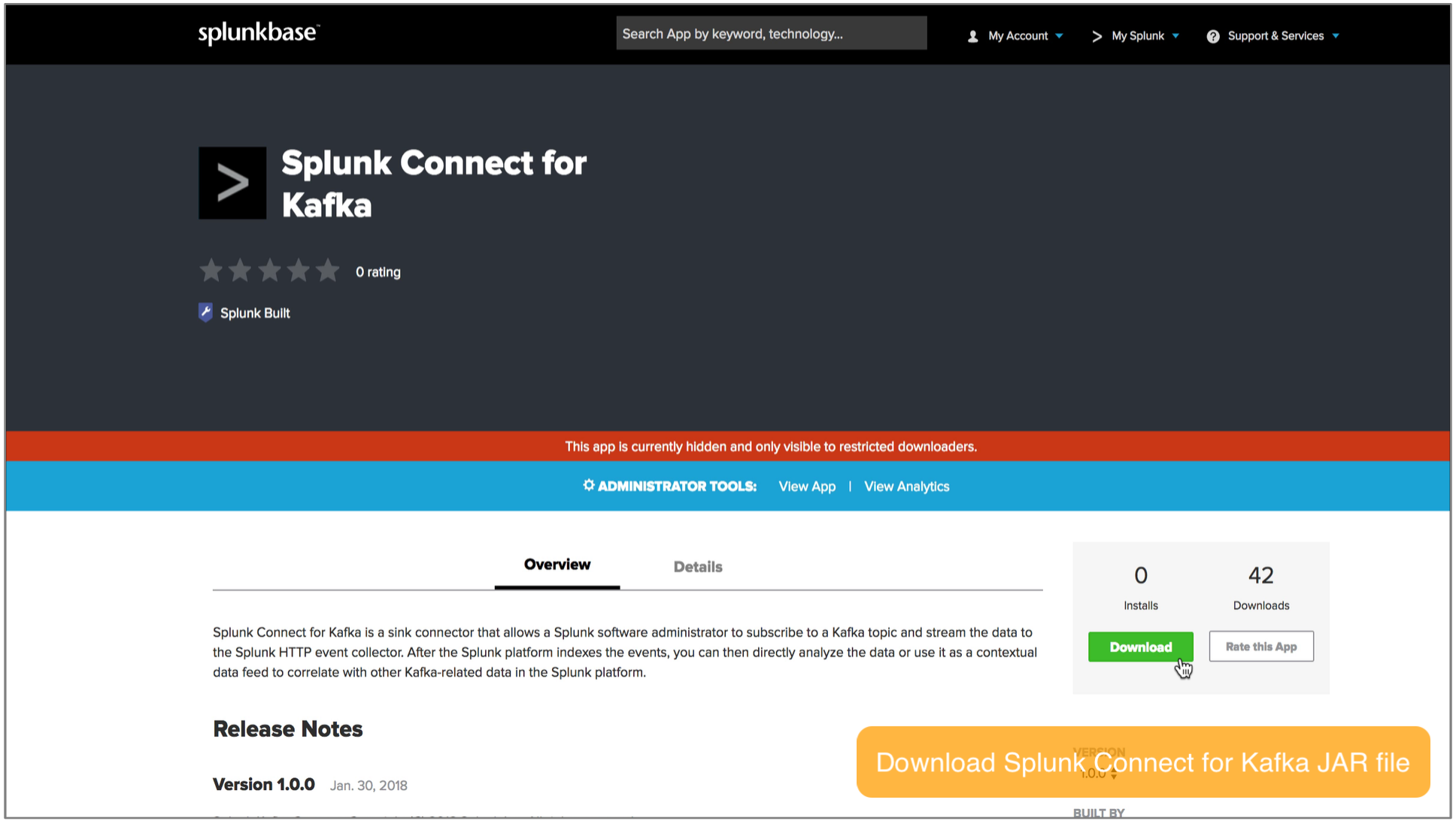
Load Balancer

- Load balancing can be accomplished either by using a dedicated load balancer, or by specifying the list of indexers in `splunk.hec.uri` when configuring the connector
- If using a load balancer, **sticky sessions** should be enabled on the load balancer, otherwise the connector won't be able to correctly track acknowledgements, blocking the send queue
- Splunk Cloud customers need to open a support case to provision an ELB that supports sticky sessions, if they have not already done so for using the AWS Kinesis Firehose integration

Set up the Connector

1. Download and Install JAR File
2. Copy File to Connectors Directory
3. Modify Kafka Connect Properties
4. Start Kafka Connect
5. Create/Deploy the Splunk Sink Connector

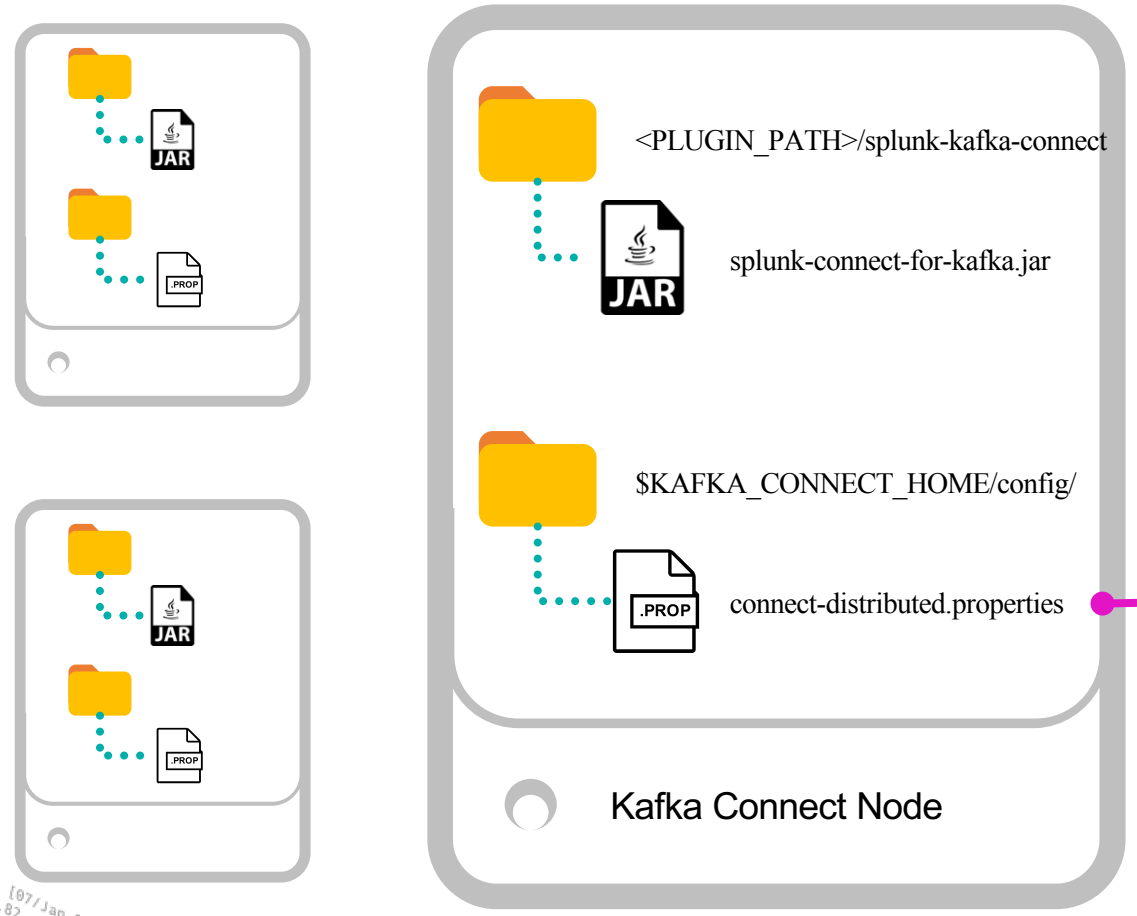
Download JAR file from Splunkbase



splunk> .conf18

Installation and Configuration

Install JAR file on each Kafka Connect node



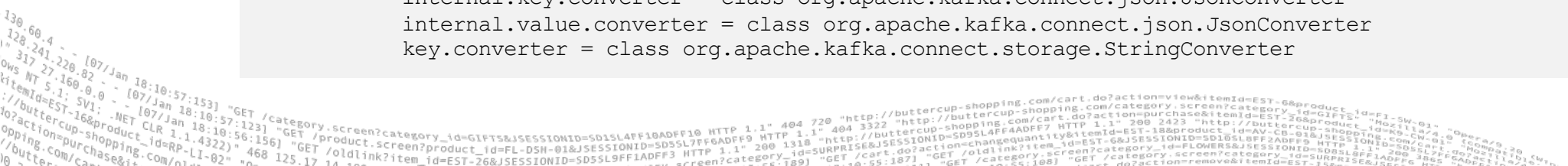
Update $KAFKA_CONNECT_HOME/config/connect-distributed.properties

```
#These settings may already be configured if you have deployed a connector in
your Kafka Connect Environment
bootstrap.servers=<BOOTSTRAP_SERVERS>
plugin.path=<PLUGIN_PATH>

#Required configurations for Splunk Kafka Connect
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 configurations for Splunk Kafka Connect
group.id=kafka-connect-splunk-hec-sink
config.storage.topic=__kafka-connect-splunk-task-configs
config.storage.replication.factor=3
offset.storage.topic=__kafka-connect-splunk-offsets
offset.storage.replication.factor=3
offset.storage.partitions=25
status.storage.topic=__kafka-connect-splunk-statuses
status.storage.replication.factor=3
status.storage.partitions=5
```

splunk> .conf18

Start Kafka Connect

```
kafka@ip-172-31-21-188:/opt/connectors/splunk-kafka-connect$ $KAFKA_CONNECT_HOME/bin/connect-distributed.sh
$KAFKA_CONNECT_HOME/config/connect-distributed.properties
[2018-03-16 06:17:37,747] INFO Kafka Connect distributed worker initializing ...
(org.apache.kafka.connect.cli.ConnectDistributed:60)
…
[2018-03-16 06:17:37,756] INFO Scanning for plugin classes. This might take a moment ...
(org.apache.kafka.connect.cli.ConnectDistributed:69)
[2018-03-16 06:17:37,765] INFO Loading plugin from: /opt/connectors/splunk-kafka-connect
(org.apache.kafka.connect.runtime.isolation.DelegatingClassLoader:184)
[2018-03-16 06:17:38,081] INFO Registered loader: PluginClassLoader{pluginLocation=file:/opt/connectors/splunk-
kafka-connect/} (org.apache.kafka.connect.runtime.isolation.DelegatingClassLoader:207)
[2018-03-16 06:17:38,082] INFO Added plugin 'com.splunk.kafka.connect.SplunkSinkConnector'
(org.apache.kafka.connect.runtime.isolation.DelegatingClassLoader:136)
[2018-03-16 06:17:38,082] INFO Added plugin 'org.apache.kafka.connect.storage.StringConverter'
(org.apache.kafka.connect.runtime.isolation.DelegatingClassLoader:136)
...
[2018-03-16 06:17:39,609] INFO Added aliases 'SplunkSinkConnector' and 'SplunkSink' to plugin
'com.splunk.kafka.connect.SplunkSinkConnector'
…
            access.control.allow.methods =
            access.control.allow.origin =
            bootstrap.servers = [localhost:9092]
            client.id =
            config.storage.replication.factor = 1
            config.storage.topic = connect-configs
            connections.max.idle.ms = 540000
            group.id = connect-cluster
            heartbeat.interval.ms = 3000
            internal.key.converter = class org.apache.kafka.connect.json.JsonConverter
            internal.value.converter = class org.apache.kafka.connect.json.JsonConverter
            key.converter = class org.apache.kafka.connect.storage.StringConverter
```

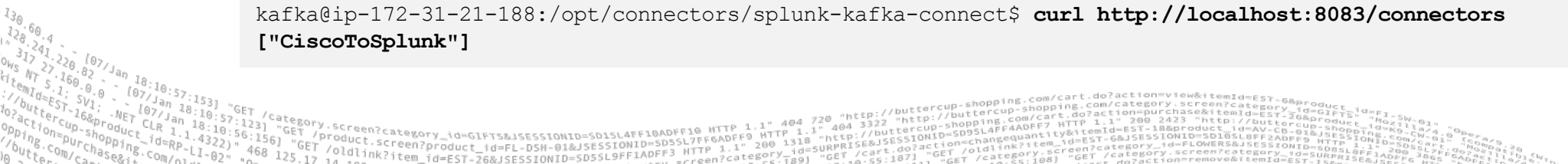


splunk> .conf18

Create instance of SplunkSinkConnector

Check that SplunkSinkConnector is an available plugin, and create connector tasks. Adjust topics to set the topic, and splunk.hec.uri and splunk.hec.token to refer to your HEC endpoint and token

```
kafka@ip-172-31-21-188:/opt/connectors/splunk-kafka-connect$ curl http://localhost:8083/connector-plugins
[{"class":"com.splunk.kafka.connect.SplunkSinkConnector","type":"sink","version":"v1.0.0"},{"class":"org.apache.k
afka.connect.file.FileStreamSinkConnector","type":"sink","version":"1.0.1"},{"class":"org.apache.kafka.connect.fi
le.FileStreamSourceConnector","type":"source","version":"1.0.1"}]

kafka@ip-172-31-21-188:/opt/connectors/splunk-kafka-connect$ curl localhost:8083/connectors -X POST -H "Content-
Type: application/json" -d '{
> "name": "CiscoToSplunk",
> "config": {
>    "connector.class": "com.splunk.kafka.connect.SplunkSinkConnector",
>    "tasks.max": "1",
>    "topics":"cisco_asa",
>    "splunk.hec.uri": "https://localhost:8088",
>    "splunk.hec.token": "9948d956-b83e-48f8-811c-f332f0082d47",
>    "splunk.hec.ack.enabled": "false",
>    "splunk.hec.ssl.validate.certs": "false"
>   }
> }'
{"name":"CiscoToSplunk","config":{"connector.class":"com.splunk.kafka.connect.SplunkSinkConnector","tasks.max":"1
","topics":"cisco_asa","splunk.hec.uri":"https://localhost:8088","splunk.hec.token":"9948d956-b83e-48f8-811c-
f332f0082d47","splunk.hec.ack.enabled":"false","splunk.hec.ssl.validate.certs":"false","name":"CiscoToSplunk"},"t
asks":[],"type":null}

kafka@ip-172-31-21-188:/opt/connectors/splunk-kafka-connect$ curl http://localhost:8083/connectors
["CiscoToSplunk"]
```

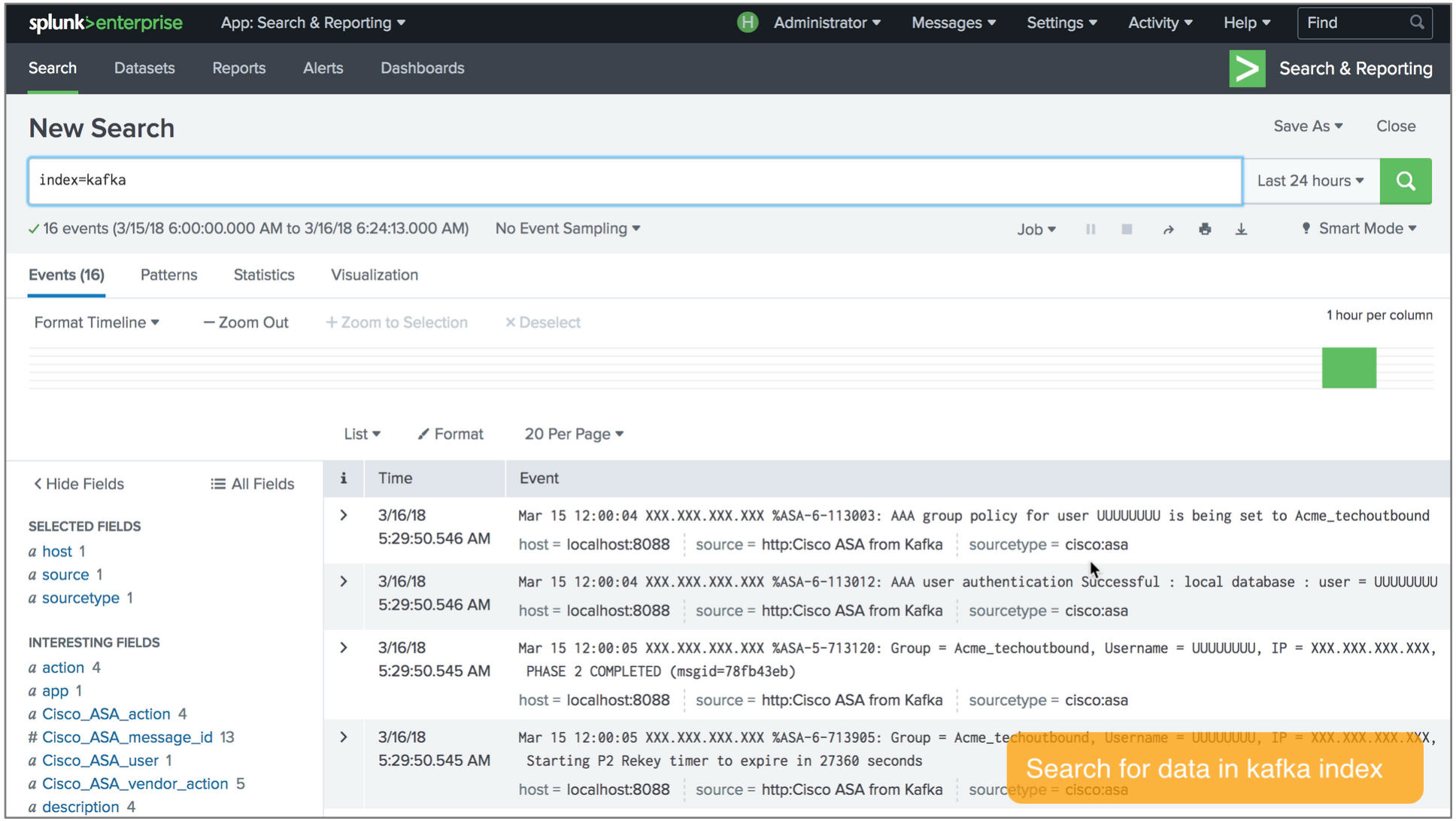


splunk> .conf18

Verify Data in Splunk

1. Search for Data in Splunk

Search for Data in Splunk



splunk> .conf18