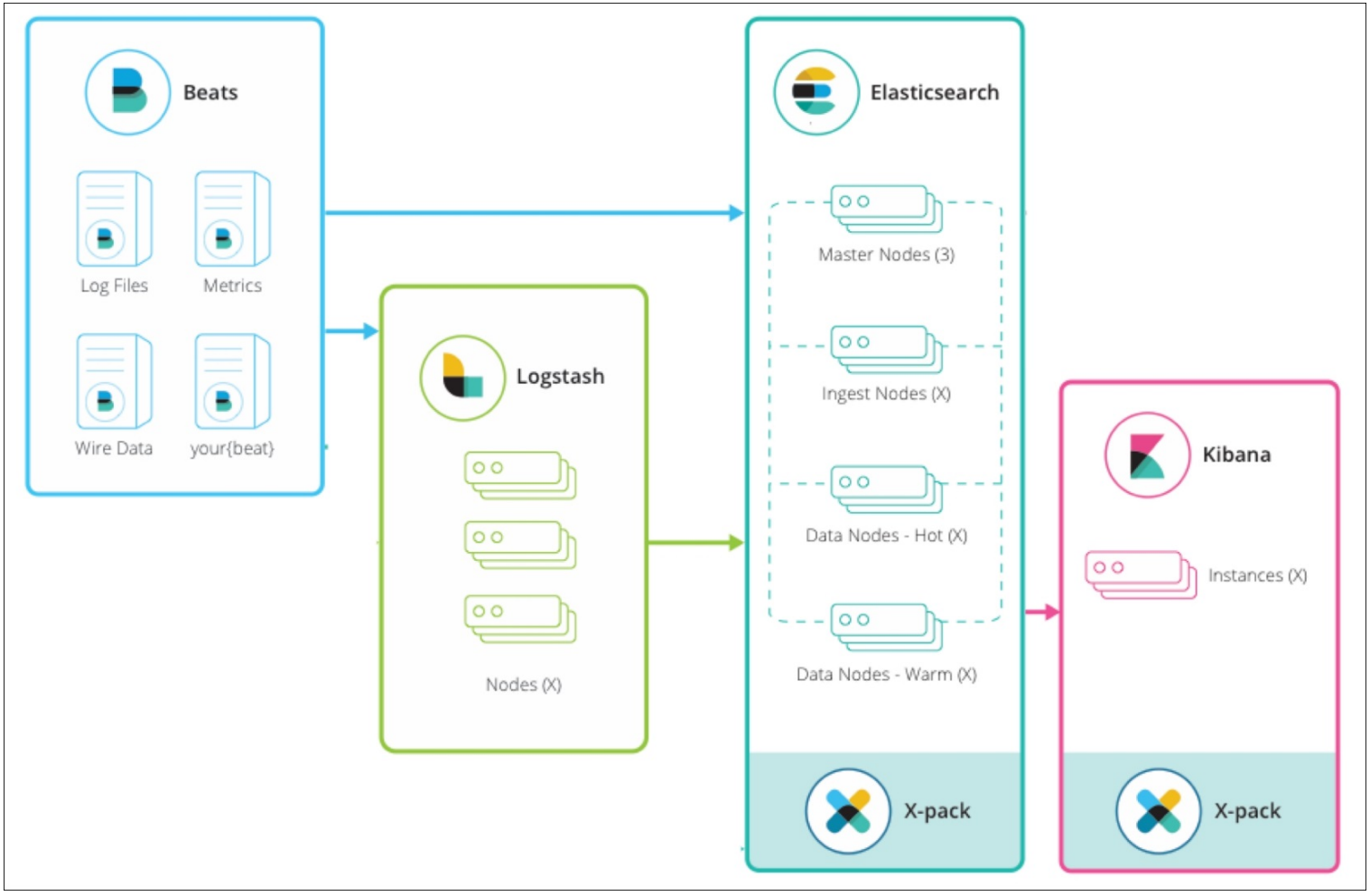
The Elastic Stack

Musa Baloyi

August 14, 2018

Table of contents

The Elastic Stack



Installation guidelines

- Installing from source
- Using a package manager
- sebp/elk docker container
- elastic/stack-docker docker container
- Elastic Team SIT
- Elastic Team SLAM

Elastic Stack demo

- sudo docker pull sebp/elk
- sudo sysctl -w vm.max_map_count=300000
- sudo docker run -p 5601:5601 -p 9200:9200 -p 5044:5044 -it –name elk sebp/elk
- Elasticsearch is running on http://localhost:9200
- Kibana is running on http://localhost:5601
- Logstash started at 5044

Elasticsearch demo

- curl -XGET 'localhost:9200/_cat/health?v&pretty'
- curl -XGET 'localhost:9200/_cat/nodes?v&pretty'
- curl -XGET 'localhost:9200/_cat/indices?v&pretty'
- Create index

```
musa@musa-VirtualBox:~$ curl -XPUT 'localhost:9200/rta-all-models2?pretty' -H 'Content-Type: application/json' -d
{
  "mappings": {
    "log": {
      "properties": {
        "env_name": {"type": "text"},
        "submit_status": {"type": "text"},
        "model_name": {"type": "text"},
        "tot_runtime": {"type": "float"}
      }
    }
  }
}'
{
  "acknowledged" : true,
  "shards_acknowledged" : true,
  "index" : "rta-all-models2"
}
```

Elasticsearch demo

- curl -XGET 'localhost:9200/_cat/indices?v&pretty'

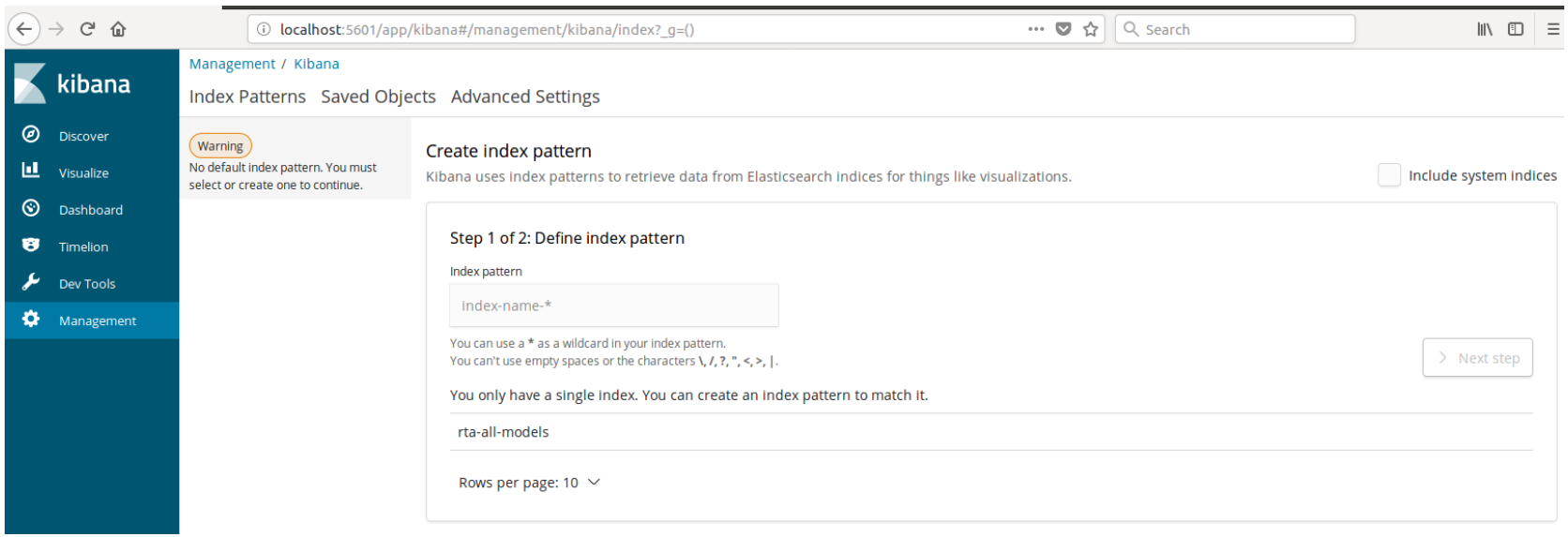
```
musa@musa-VirtualBox:~$ curl -XGET 'localhost:9200/_cat/indices?v&pretty'
health status index               uuid                   pri rep docs.count docs.deleted store.size pri.store.size
yellow open   logstash-2015.05.18 sKtfBC7HSMWBBgfUcPUidg   5   1          0            0      1.1kb          1.1kb
yellow open   rta-all-models      MkgSfii6QrytEepm6uBw4w   5   1          0            0      1.1kb          1.1kb
yellow open   rta-all-models2     lIpB3IE7QhOFRom6UJsYqg   5   1          0            0      1.1kb          1.1kb
```

- curl -H 'Content-Type: application/x-ndjson' -XPOST 'localhost:9200/_bulk?pretty' –data-binary @rta_2018-03-13T08 37 47.143254.json

Elasticsearch clients

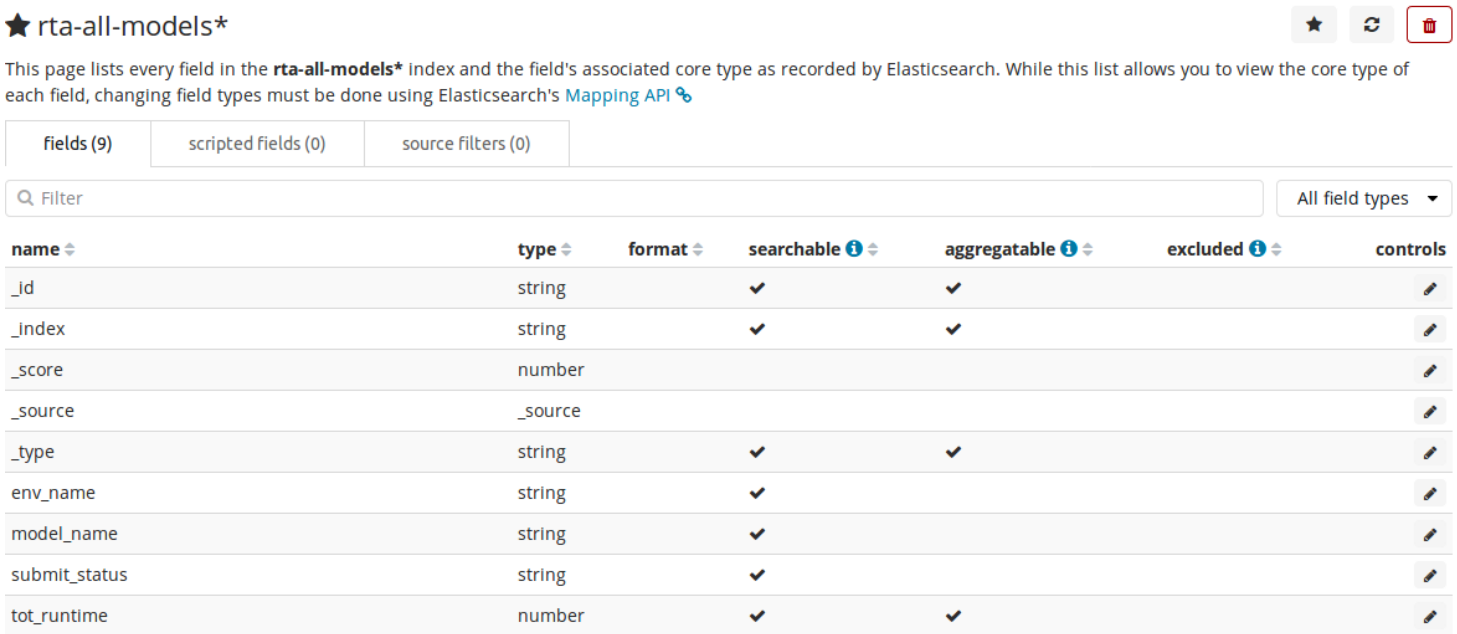
Kibana demo

- Access Kibana to see created indices and data



Kibana demo

- Create index pattern in Kibana



★ rta-all-models*

This page lists every field in the **rta-all-models*** index and the field's associated core type as recorded by Elasticsearch. While this list allows you to view the core type of each field, changing field types must be done using Elasticsearch's Mapping API

fields (9) | scripted fields (0) | source filters (0)

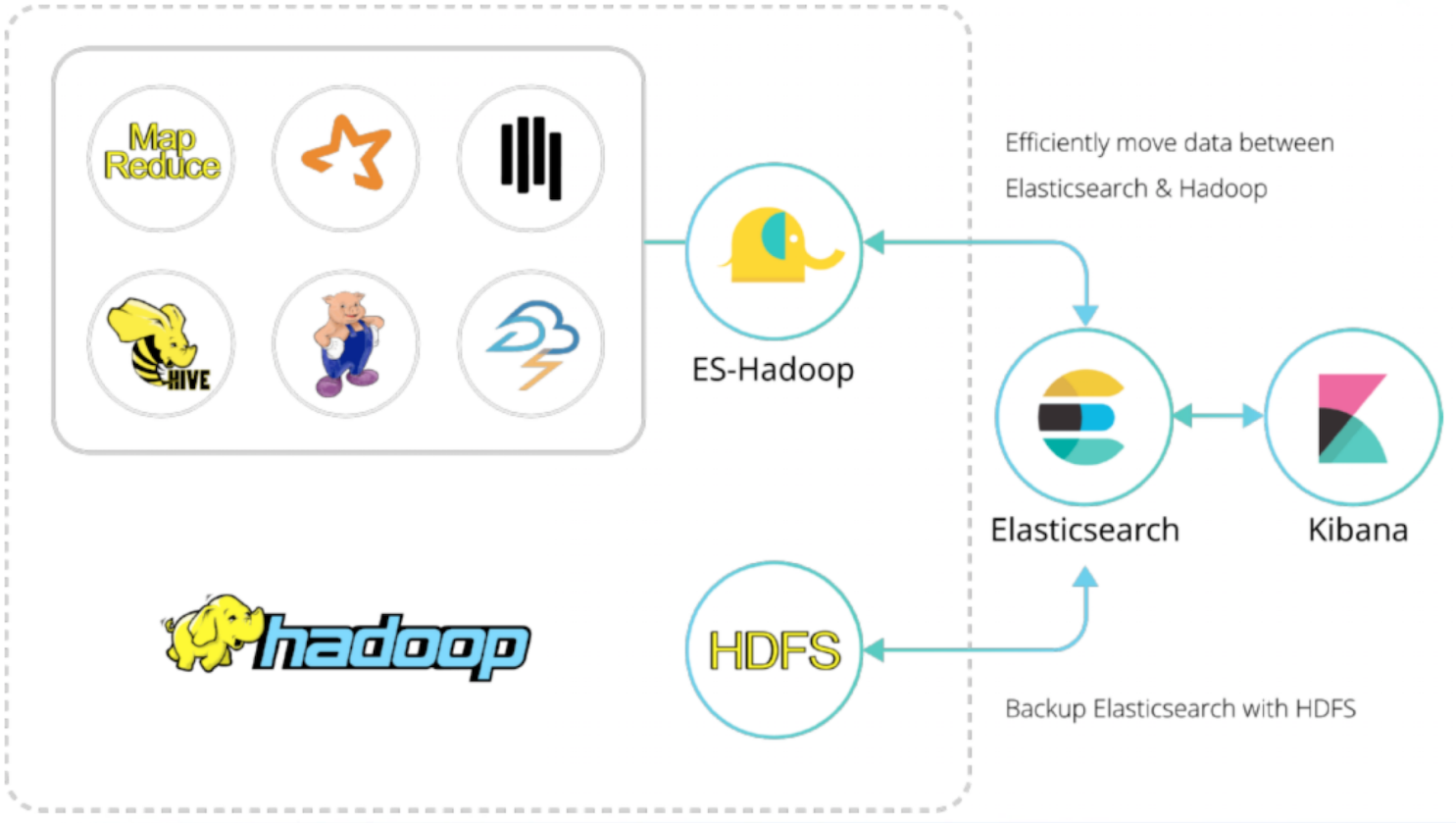
Filter | All field types

name	type	format	searchable	aggregatable	excluded	controls
_id	string		✔	✔		
_index	string		✔	✔		
_score	number					
_source	_source					
_type	string		✔	✔		
env_name	string		✔			
model_name	string		✔			
submit_status	string		✔			
tot_runtime	number		✔	✔		

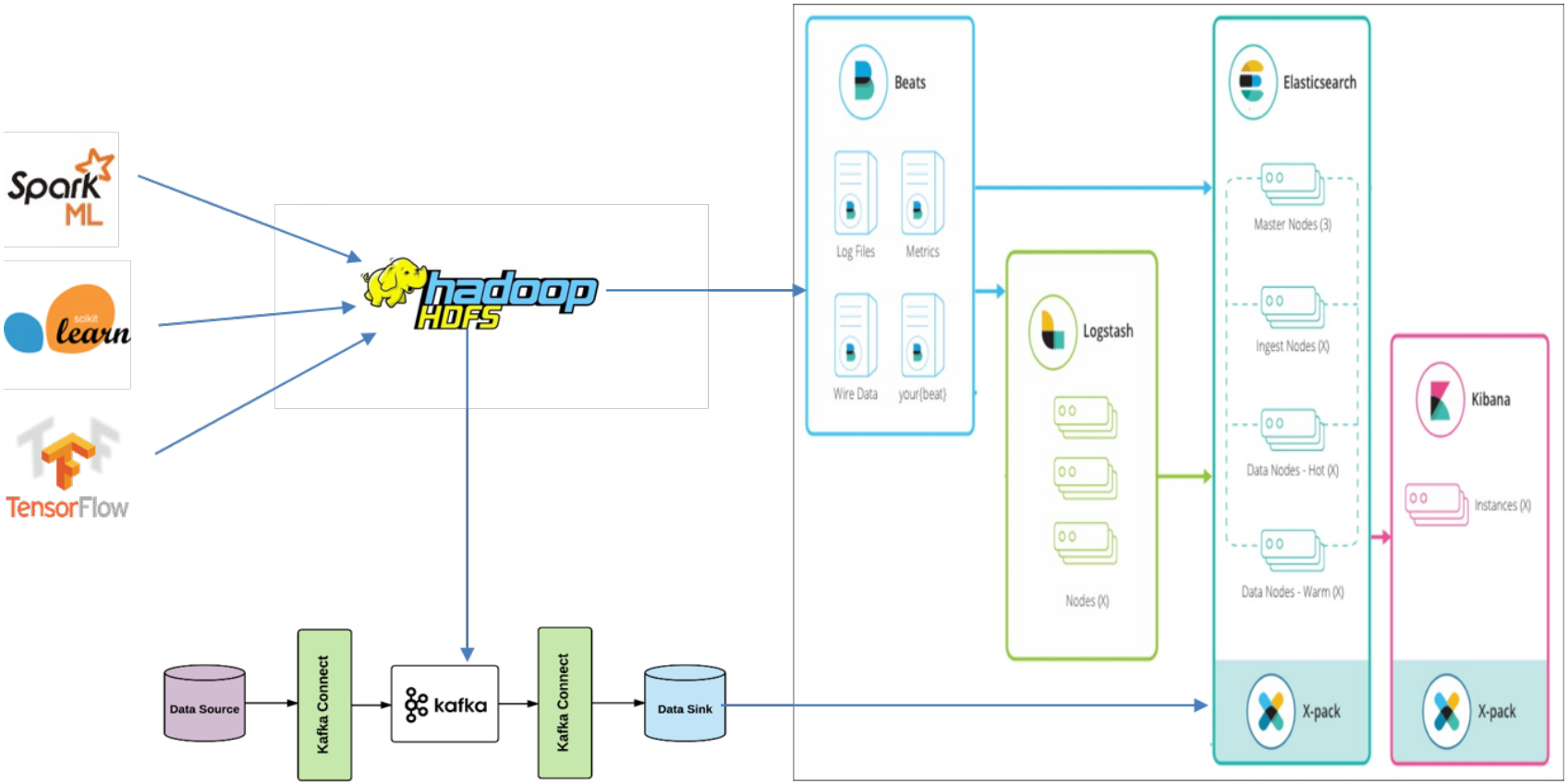
Kibana Console

- Alternative to CURL

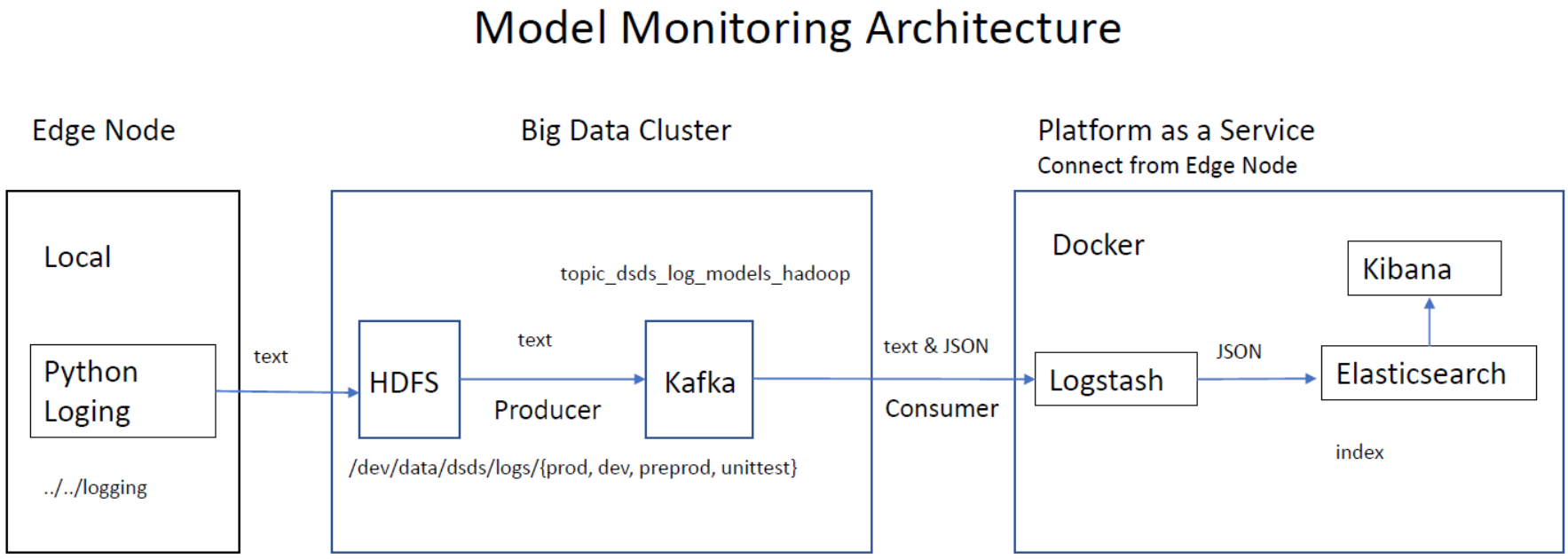
Elasticsearch for Hadoop



Architectural overview



Architectural overview



Kafka

- Kafka is generally used for building real-time streaming
  - data pipelines that reliably get data between systems or applications
  - applications that transform or react to the streams of data
- Kafka is run as a cluster on one or more servers that can span multiple datacenters.
- The Kafka cluster stores streams of records in categories called topics.
- Each record consists of a key, a value, and a timestamp.

Kafka installation

- Download the binary: kafka_2.12-1.0.1.tgz
- 7z x kafka_2.12-1.0.1.tgz && 7z x kafka_2.12-1.0.1.tar
- sudo mv kafka_2.12-1.0.1 /opt/Kafka

Kafka demo

- cd /opt/Kafka/ kafka_2.12-1.0.1
- sudo bin/kafka-server-start.sh config/server.properties
- bin/kafka-console-consumer.sh –bootstrap-server localhost:9092 –topic testing –from-beginning
- bin/kafka-topics.sh –create –zookeeper localhost:2181 –replication-factor 1 –partitions 1 –topic testing
- bin/kafka-topics.sh –list –zookeeper localhost:2181
- Configure Kafka producer connect-file-source.properties
- Configure Kafka consumer connect-file-sink.properties
- bin/connect-standalone.sh config/connect-standalone.properties config/connect-file-source.properties config/connect-file-sink.properties

Beats

- Seamlessly integrates with the Elastic Stack. Kafka requires a separate install.
- One way, Kafka is bidirectional
- Extensible
- Shippers: Filebeat, Metricbeat, Packetbeat, Winlogbeat, Auditbeat, Heartbeat.
- Filebeat configuration

Beats demo: Heartbeat

- elastic/stack-docker

Logstash.conf

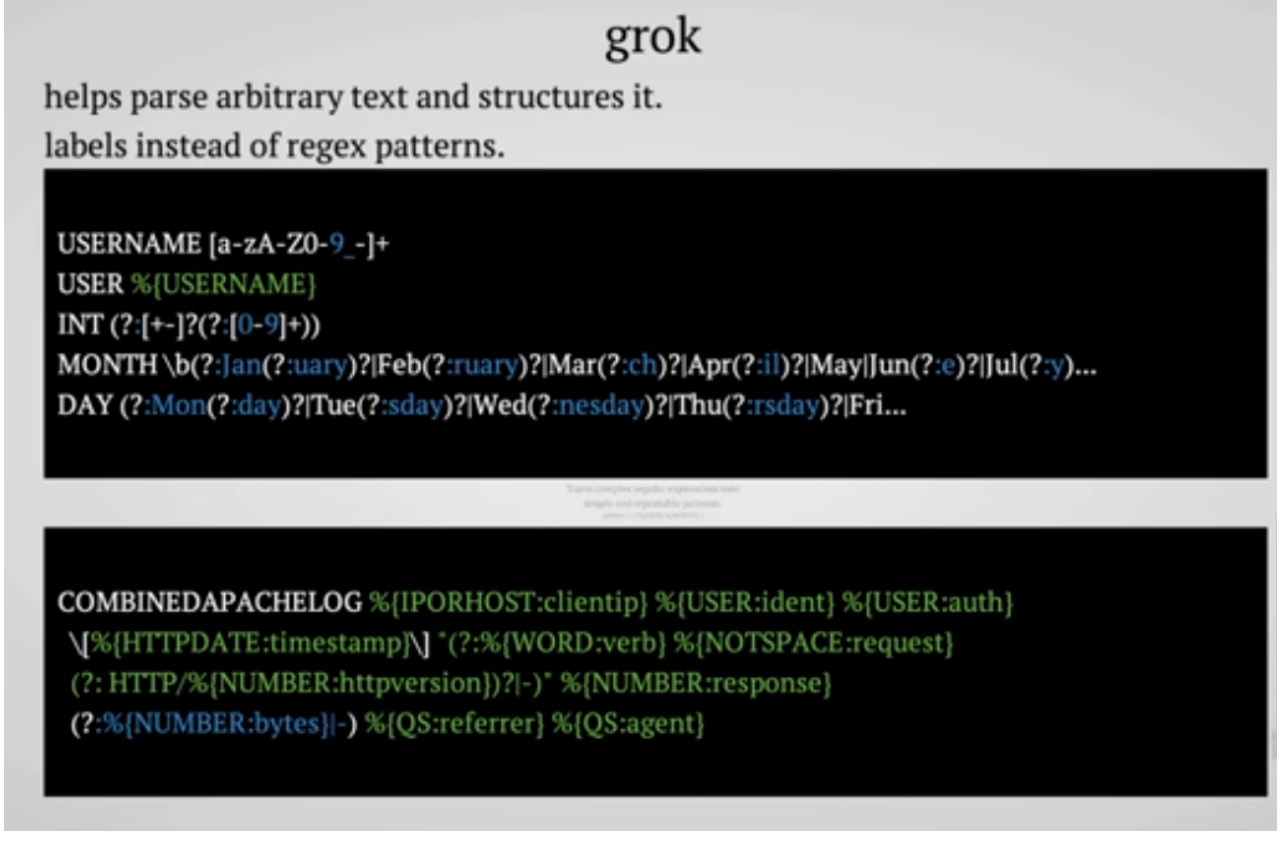
```
input {
  heartbeat {
    interval => 5
    message  => 'Hello from Logstash 🐭'
  }
}

#input {
#   filebeat {
#        port => 5044
#   }
#}

output {
  elasticsearch {
    hosts    => [ 'elasticsearch' ]
    user     => 'elastic'
    password => 'changeme'
  }
}
```

Logstash

- grok – better pattern matching
- https://www.elastic.co/guide/en/logstash/current/plugins-filters-grok.html



References

1. Building Intelligent Systems: A Guide to Machine Learning Engineering. [Geoff Hulten] (Apress, 2018)
2. Trends in AI, Data Science, and Big Data. [Ben Lorica] (2017)
3. Building Evolutionary Architectures. [Rebecca Parsons; Patrick Kua; Neal Ford] (O'Reilly Media, 2017)
4. 5 things you should be monitoring. [Brian Brazil] (2018)
5. The Logstash Book. [James Turnbull] (Turnbull Press, 2013)
6. Beyond the Twelve-Factor App. [Kevin Hoffman] (O'Reilly Media, 2016)
7. Logs and real-time stream processing. [Jay Kreps] (2016)
8. I Heart Logs: Apache Kafka and Real-time Data Integration. [Jay Kreps] (2015)
9. The log: The lifeblood of your data pipeline. [Kiyoto Tamura] (2015)
10. Understanding the ELK stack. [Brian Anderson; Rafał Kuć] (2016)