Elastic (ELK) Stack

elasticsearch logstash kibana

Agenda

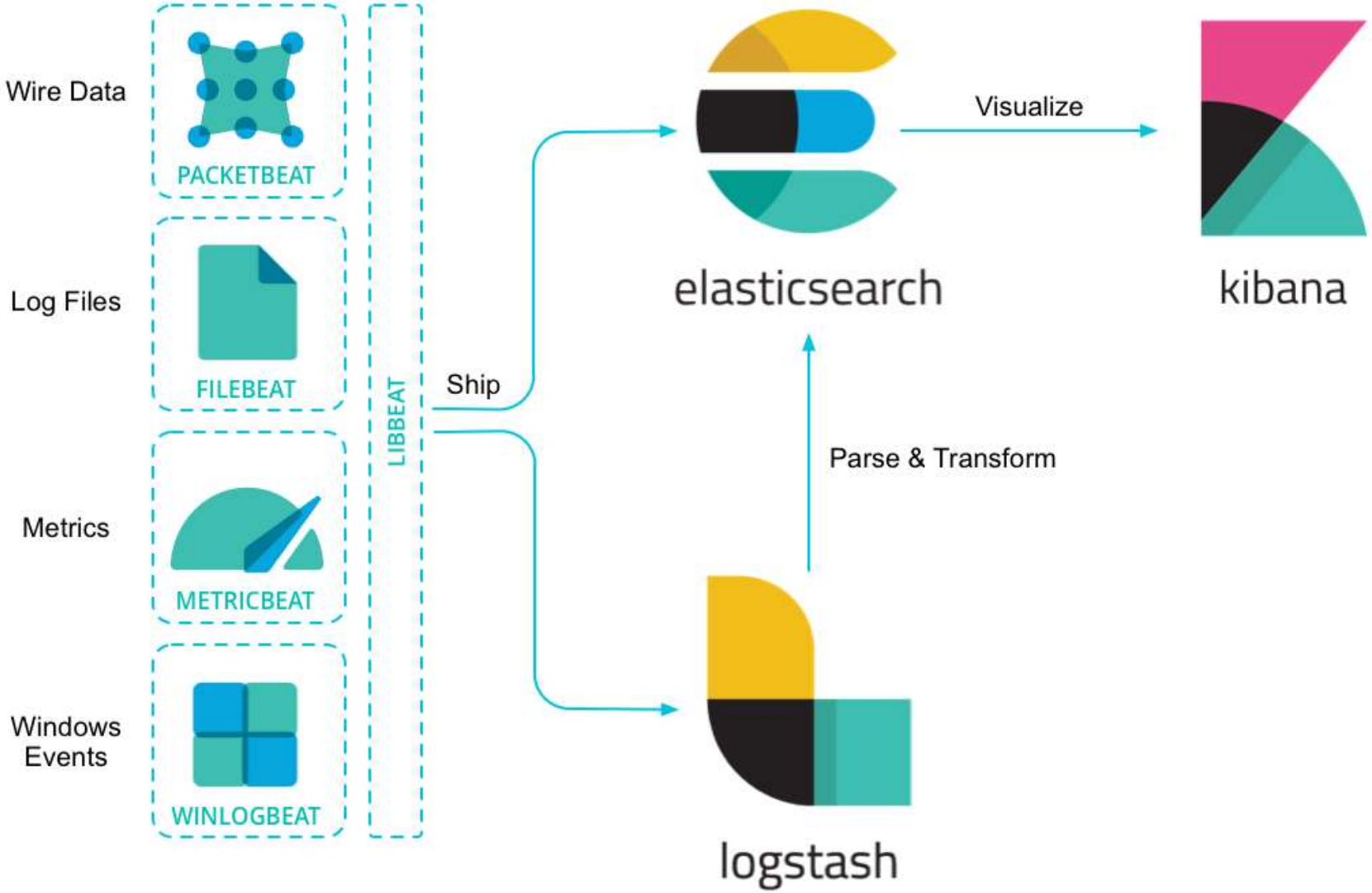
- Introduction
- Elastic Stack Overview
- Components of Elastic Stack
- Role of Elastic Stack in Big Data Analysis
- Demo
  - ElasticSearch configurations
  - Logstash pipelines
  - Kibana Dashboards
  - Beats example
  - Twitter trend example
- Q & A

Elastic (ELK) Stack

Elastic Stack is a group of open source products from Elastic designed to help users take data from any type of source and in any format and search, analyze, and visualize that data in real time. It uses Logstash for log aggregation, Elasticsearch for searching, and Kibana for visualizing and analyzing data.

- **ElasticSearch:** Store, Search, and Analyze
- **Logstash:** Collect logs and events data, Parse and Transform
- **Kibana:** Explore, Visualize, and Share
- **Beats:** Data shipper.

Elastic (ELK) Stack Architecture



ElasticSearch

Elasticsearch is a highly available and distributed search engine.

- Built on top of Apache Lucene
- NoSQL Datastore
- Schema-free
- JSON Document
- RESTful APIs

Relational Database	ElasticSearch
Database	Index
Table	Type
Row	Document
Column	Field
Schema	Mapping

- Node
- Cluster

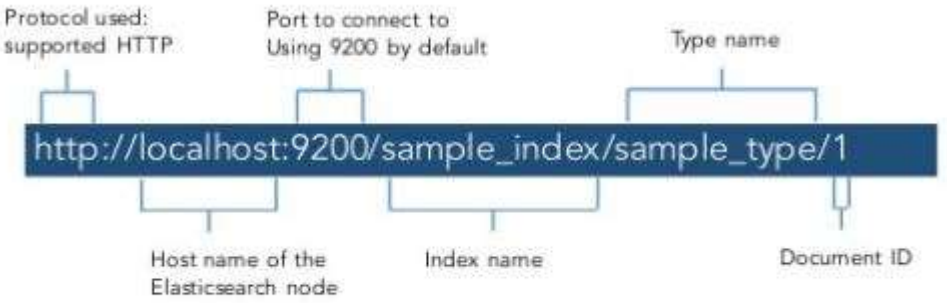
ElasticSearch

Elasticsearch is distributed, which means that indices can be divided into **shards** and each shard can have zero or more **replicas**. By default, an index is created with 5 shards and 1 replica per shard (5/1). Rebalancing and routing of shards are done automatically.

Features

- Distributed
- Scalable
- Highly available
- Near Real Time (NRT) search
- Full Text Search
- Java, .NET, PHP, Python, Curl, Perl, Ruby
- HADOOP & SPARK -- Elasticsearch-Hadoop (ES-Hadoop)

ElasticSearch RESTful API



HTTP Based CRUD Operations

Operation	CURL command
Create	curl –XPUT “http://localhost:9200/<index>/<type>/<id>”
Read	curl –XGET “http://localhost:9200/<index>/<type>/<id>”
Update	curl –XPOST “http://localhost:9200/<index>/<type>/<id>”
Delete	curl –XDELETE “http://localhost:9200/<index>/<type>/<id>”

curl -X GET 'http://localhost:9200/_cat/indices?v'

GitHub Casestudy

Challenge : How do you satisfy the search needs of GitHub's **4 million users** while simultaneously providing tactical operational insights that help you iteratively improve customer service?

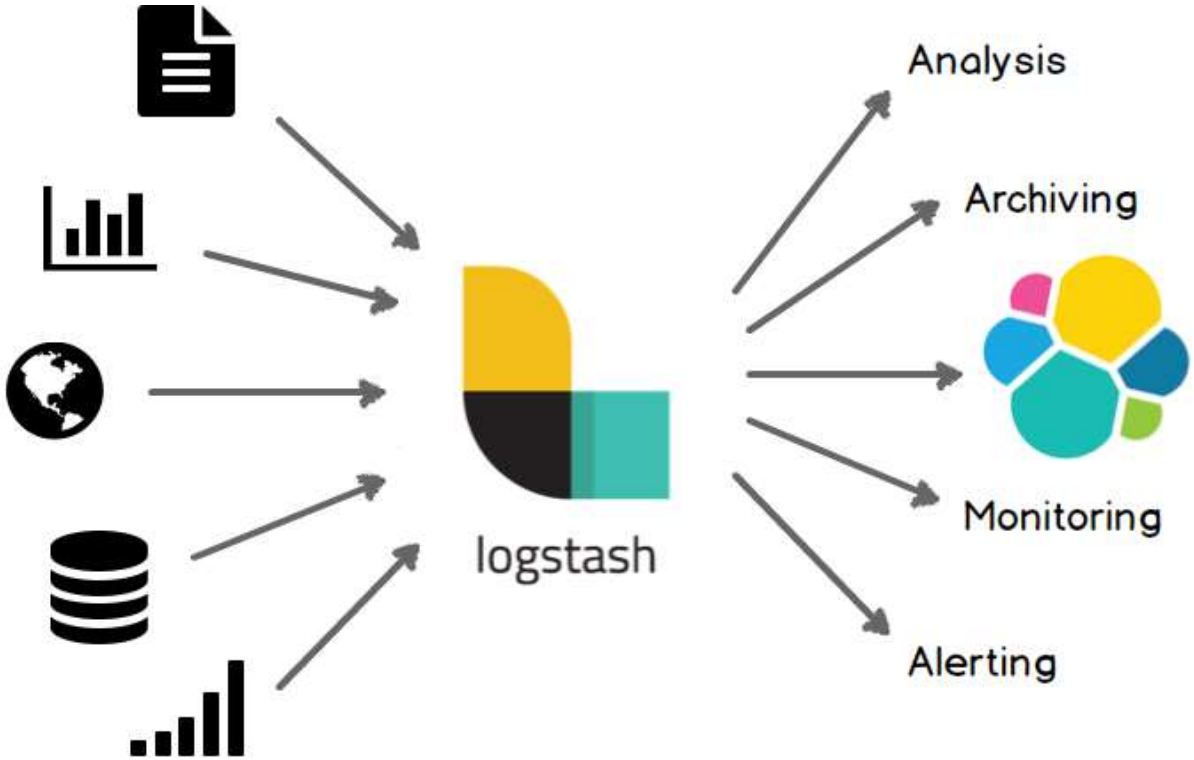
"Search is at the core of GitHub"

Solution: GitHub uses Elasticsearch to continually index the data from an ever-growing store of over **8 million code repositories**, comprising over **2 billion documents.**
GitHub uses Elasticsearch to index new code as soon as users push it to a repository on GitHub.

Other customers includes Facebook, Netflix, ebay, Wikimedia, etc.
ebay : Searching across 800 million listings in subseconds

Logstash

Logstash can collect logs from a variety of sources (using **input plugins**), process the data into a common format using **filters**, and stream data to a variety of sources (using **output plugins**). Multiple filters can be chained to parse the data into a common format. Together, they build a [Logstash Processing Pipeline](#).



Logstash Plug-ins

Logstash has a rich collections of input, filter and output plugins. You can now create your own Logstash plugin and add it into community plugins.

Input Plugins

- Beats
- Elasticsearch
- File
- Graphite
- Heartbeat
- Tttp
- Jdbc
- Kafka
- Log4j
- Redis
- Stdin
- TCP
- Twitter

Filter Plugins

- Aggregate
- csv
- Date
- geoip
- Grok
- Json
- sleep
- urlencode
- UUID
- xml

Output Plugins

- CSV
- Elasticsearch
- Email
- File
- Graphite
- Http
- Jira
- Kafka
- Nagios
- Redis
- Stdout
- S3
- Tcp
- Udp

Logstash Pipeline

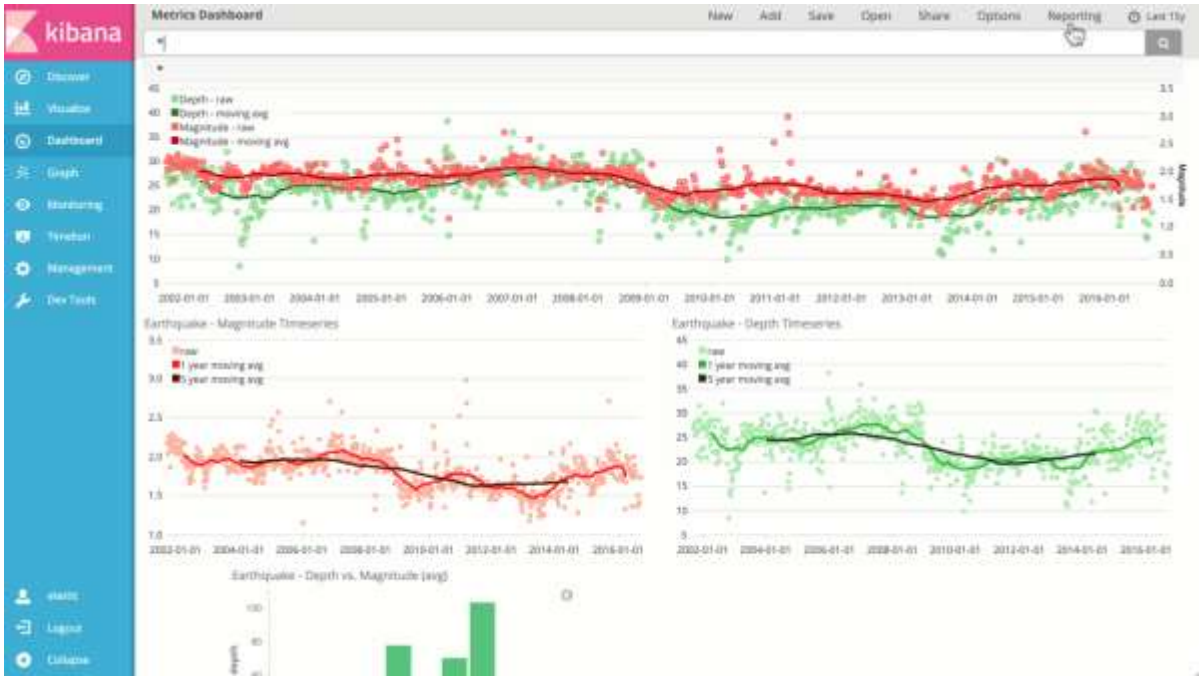
Basic Configuration of Logstash Pipeline

```
input {
  file {
    path => "/tmp/access_log"
    start_position => "beginning"
  }
}

filter {
  if [path] =~ "access" {
    mutate { replace => { "type" => "apache_access" } }
    grok {
      match => { "message" => "%{COMBINEDAPACHELOG}" }
    }
  }
  date {
    match => [ "timestamp" , "dd/MMM/yyyy:HH:mm:ss Z" ]
  }
}

output {
  elasticsearch {
    hosts => ["localhost:9200"]
  }
  stdout { codec => rubydebug }
}
```

Kibana

Kibana gives you the freedom to select the way you give shape to your data.

- Discover
- Visualise
- Dashboards
- Put Geo Data on Any Map
- Insert dashboards into your internal wiki or webpage
- Send your coworker a URL to a dashboard.

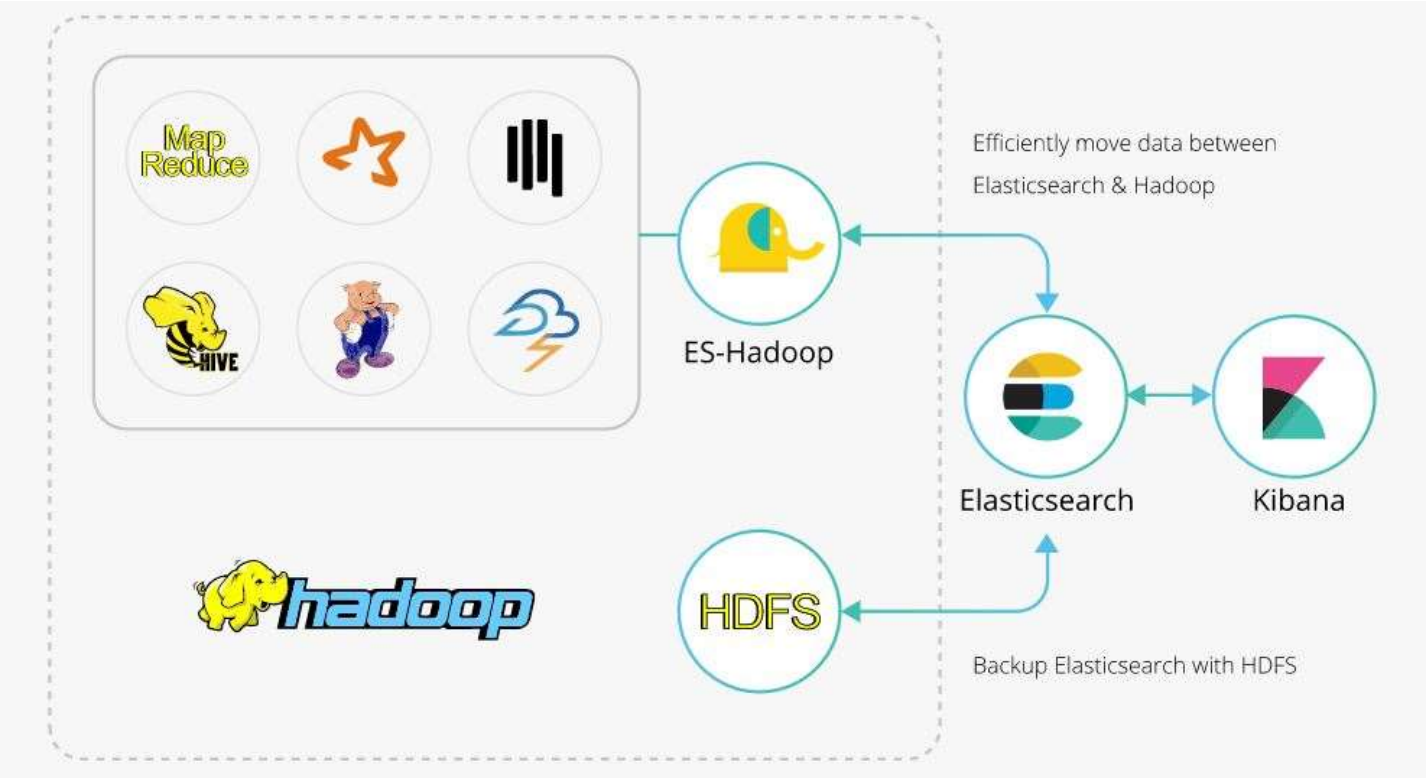


Beats

Lightweight Data Shippers.
Beats is the platform for single-purpose data shippers. They install as lightweight agents and send data from hundreds or thousands of machines to Logstash or Elasticsearch.

Elastic Stack for Big Data Analysis

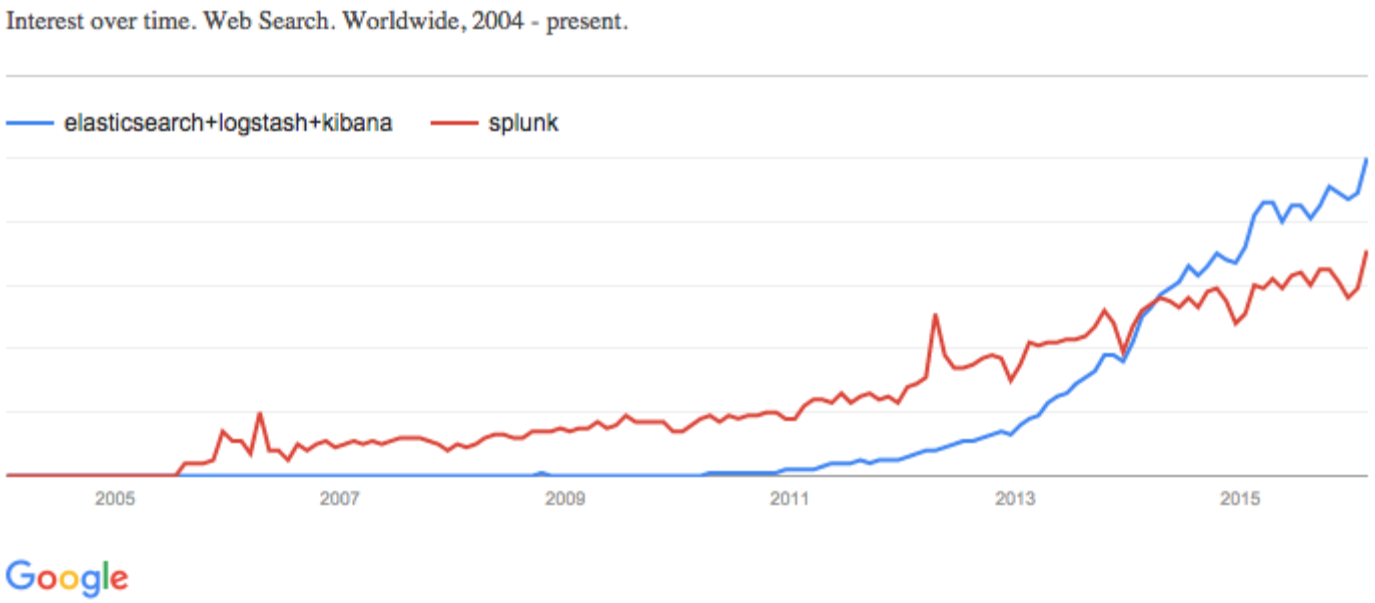
Connect the massive data storage and deep processing power of Hadoop with the real-time search and analytics of Elasticsearch.
ES-Hadoop lets you index Hadoop data into the Elastic Stack to take full advantage of the speedy Elasticsearch engine and beautiful Kibana visualizations.



Elasticsearch for Apache Hadoop

Splunk VS ELKStack

Splunk and the ELK stack are dominating the interest in the log management space with the most comprehensive and customizable solutions.



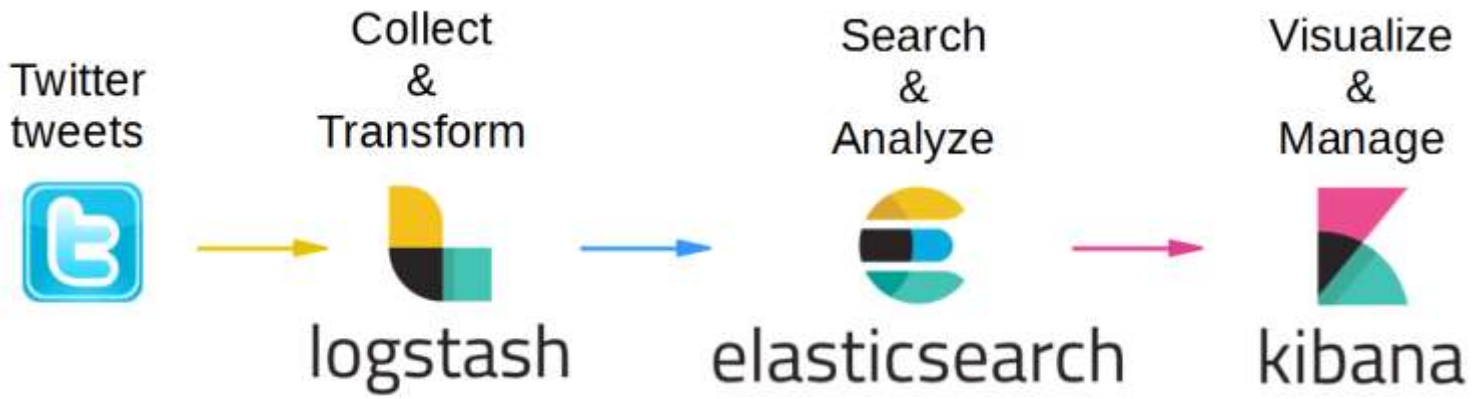
Popularity Trend

A head to head comparison is always a tough call, especially when there's no clear winner and the tool you choose can potentially have a huge impact on the business

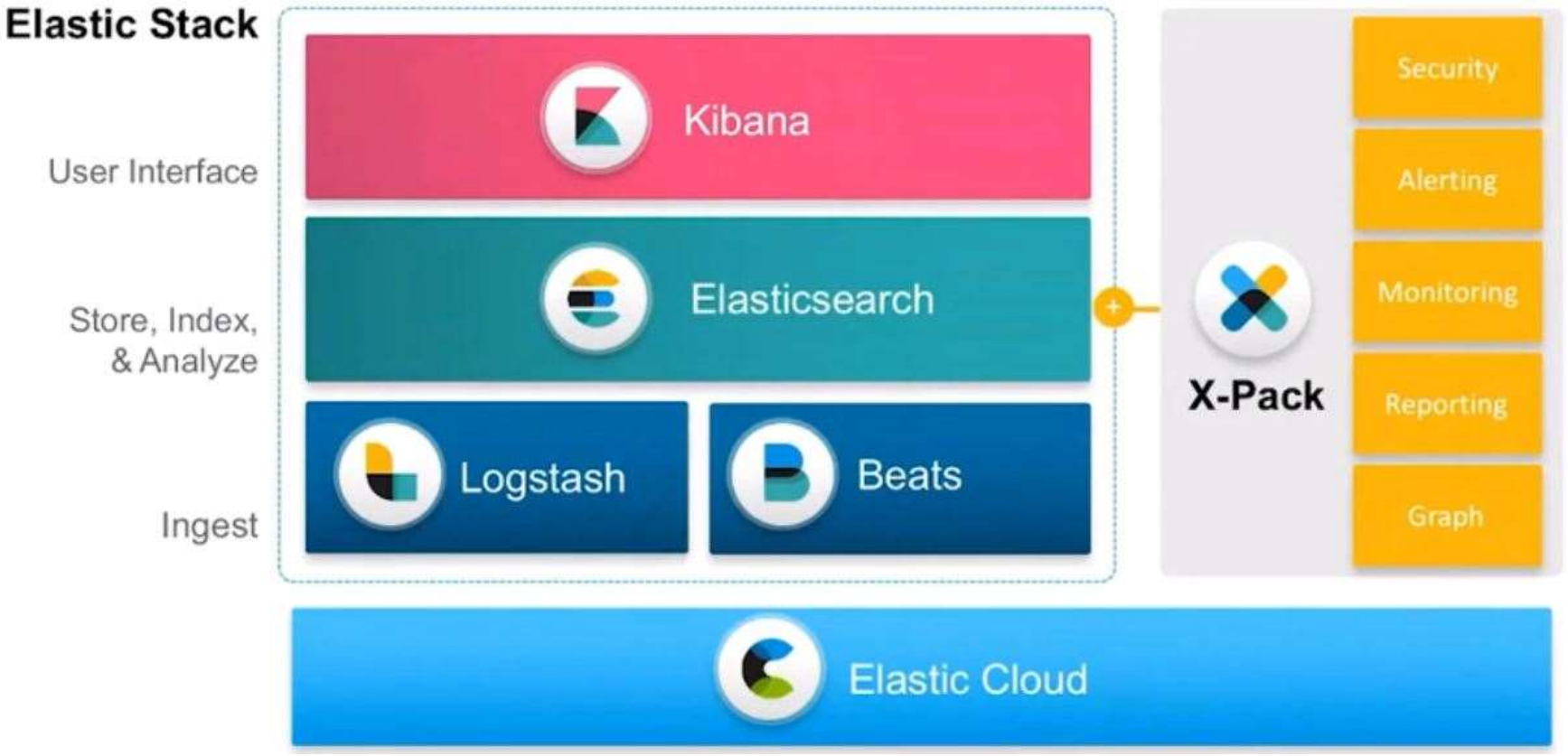
Demo !!

Basic Example

Demo – Twitter Example



X-Pack & Elastic Cloud



Summary

- Elastic Stack
- Components of Elastic Stack
- Configurations
- **ES-Hadoop** plugin for Big Data Analysis
- **ElasticSearch** : Store, Search , Analysis
- **Logstash**: ETL
- **Kibana**: Visualisation
- Beats: Data Shipper
- Elastic Cloud