ELK Stack Tutorial

What is the ELK Stack?

The **ELK Stack** is a collection of three open-source products — Elasticsearch, Logstash, and Kibana. ELK stack provides centralized data processing/storing in order to identify problems with servers or applications. It allows you to search all the data in a single place.

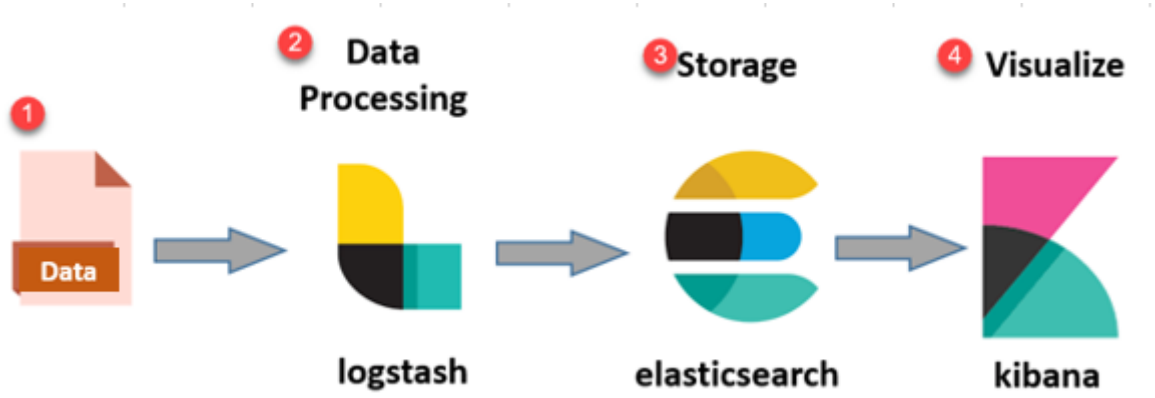
- **E** stands for ElasticSearch: used for storing data
- **L** stands for LogStash : used for both shipping as well as processing and storing data
- **K** stands for Kibana: is a visualization tool (a web interface) which is hosted through Nginx or Apache

ElasticSearch, LogStash and Kibana are all developed, managed ,and maintained by the company named Elastic.

ELK Stack is designed to allow users to take data from any source, in any format, and to search, analyze, and visualize that data in real time.

ELK Stack Architecture

Here is the simple architecture of ELK stack



ELK Stack Architecture

- **Data:** Server data that need to be analyzed are identified
- **Logstash:** Collect data and events data. It even parses and transforms data
- **ElasticSearch:** The transformed data from Logstash is Store, Search, and indexed.
- **Kibana:** Kibana uses Elasticsearch DB to Explore, Visualize, and Share

However, one more component is needed or Data collection called File Beats/FSCrawler. This led Elastic to rename ELK as the Elastic Stack.

What is Elasticsearch?

– Document (Json) oriented search engine
– Distributed
– Horizontally scalable and Highly Available
– Multi-tenancy enabled
– API centric & RESTful
– Built on Lucene search engine library

& used for

– full-text search, structured search, analytics, or all three in combination

Advantages of Elasticsearch

- Store schema-less data and also creates a schema for your data
- Perform filtering and querying your data for insights
- Based on Apache Lucene and provides RESTful API
- Provides horizontal scalability, reliability, and multitenant capability for real time use of indexing to make it faster search

Important Terms used in Elastic Search

Term	Description (~analogy with relational database)
Cluster	~Database cluster Group of nodes
Node	~Instance of database A JVM process, usually a machine
Index	~Database schema Hosts mapping types and their definitions contains many shards
Mapping Type	~Database Table Field description, indexing requirements
Document	~Database row Json document.
Shard	A Lucene index. Scalable unit and heart of search engine (*primary and replica)*

Document Metadata Fields

- ***_id*** - The id of the document
- ***_type*** - The document type
- ***_source*** - enabled Stores the original document that was indexed
- ***_all*** enabled Indexes all values of all document fields
- ***_timestamp*** disabled timestamp associated with the document
- ***_ttl*** disabled optionally defines an expiration time
- ***_size*** disabled indexes the size of the uncompressed

What is Logstash?

Logstash is the data collection pipeline tool. It collects data inputs and feeds into the Elasticsearch. It gathers all types of data from the different source and makes it available for further use.

Logstash can unify data from disparate sources and normalize the data into your desired destinations. It allows you to cleanse and democratize all your data for analytics and visualization of use cases.

It consists of three components:

- **Input**: passing data to process them into machine understandable format
- **Filters**: It is a set of conditions to perform a particular action or event
- **Output**: Decision maker for processed event or log

Features of Logstash

- Events are passed through each phase using internal queues
- Allows different inputs for your data
- Filtering/parsing for your data

Advantage of Logstash

- Offers centralize the data processing
- It analyzes a large variety of structured/unstructured data and events
- ELK LogStash offers plugins to connect with various types of input sources and platforms

What is Kibana?

Kibana is a data visualization which completes the ELK stack. This tool is used for visualizing the Elasticsearch documents and helps developers to have a quick insight into it. Kibana dashboard offers various interactive diagrams, geospatial data, and graphs to visualize complex quires.

It can be used for search, view, and interact with data stored in Elasticsearch directories. Kibana helps you to perform advanced data analysis and visualize your data in a variety of tables, charts, and maps.

In Kibana there are different methods for performing searches on your data.

Here are the most common search types:

Search Type	Usage
Free text searches	It is used for searching a specific string
Field-level searches	It is used for searching for a string within a specific field
Logical statements	It is used to combine searches into a logical statement.
Proximity searches	It is used for searching terms within specific character proximity.

Now in this Kibana tutorial, let's learn about important features of Kibana:

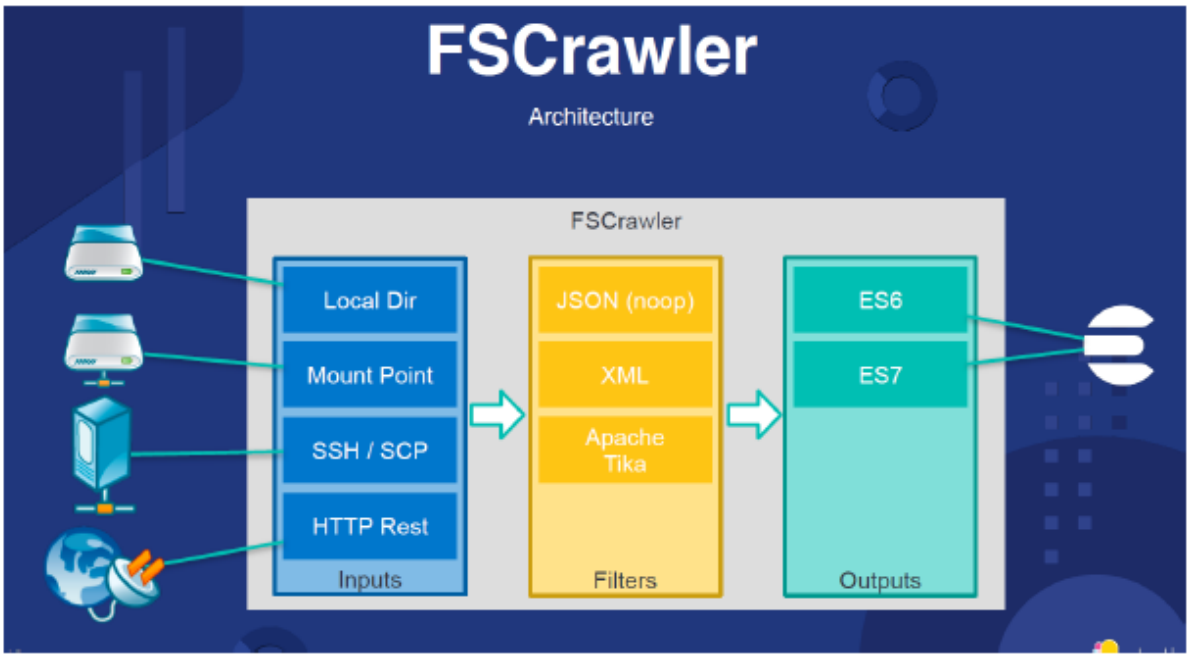
Features of Kinbana:

- Powerful front-end dashboard which is capable of visualizing indexed information from the elastic cluster
- Enables real-time search of indexed information
- You can search, View, and interact with data stored in Elasticsearch

- Execute queries on data & visualize results in charts, tables, and maps
- Configurable dashboard to slice and dice logstash data in elasticsearch
- Capable of providing historical data in the form of graphs, charts, etc.
- Real-time dashboards which is easily configurable
- Kibana ElasticSearch enables real-time search of indexed information

Advantages and Disadvantages of Kinbana

- Easy visualizing
- Fully integrated with Elasticsearch
- Visualization tool
- Offers real-time analysis, charting, summarization, and debugging capabilities
- Provides instinctive and user-friendly interface
- Allows sharing of snapshots of the data searched through
- Permits saving the dashboard and managing multiple dashboards



FSCrawler

Key Features

- Much more formats than ingest attachment plugin
- OCR (Tesseract)
- Much more metadata than ingest attachment plugin
  (See https://fscrawler.readthedocs.io/en/latest/admin/fs/elasticsearch.html#generated-fields)
- Language detection

Documentation

- https://fscrawler.readthedocs.io/
- https://fscrawler.readthedocs.io/en/latest/user/tutorial.html
- https://fscrawler.readthedocs.io/en/latest/user/formats.html
- https://fscrawler.readthedocs.io/en/latest/admin/fs/index.html