elastic stack

Introduzione

Elasticsearch è un motore di ricerca e analytics
open source basato sulla libreria **Apache Lucene**.

E' progettato per gestire grandi quantità di dati in tempo reale.

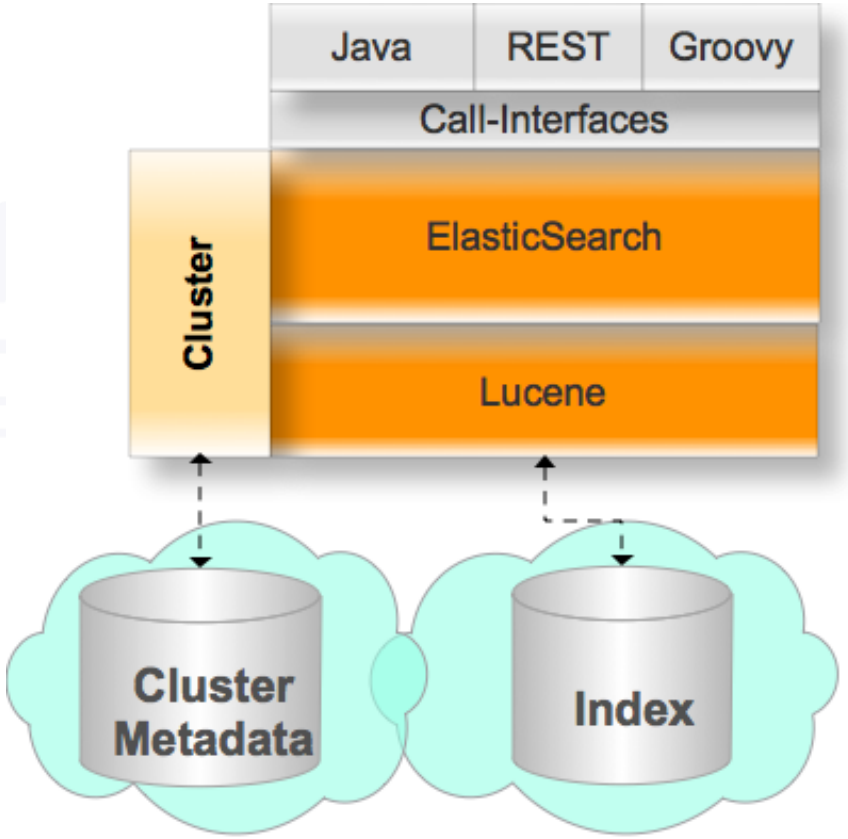
E' un database **NoSQL** orientato ai documenti e ai dati testuali.

Lucene è una libreria open-source (Java-Based) che espone, tramite le sue **API**, un motore di ricerca full-text ad alte prestazioni.

Lucene può indicizzare qualsiasi documento che sia possibile convertire in formato testuale.

Elasticsearch estende le funzionalità di Lucene con lo scopo di avere:

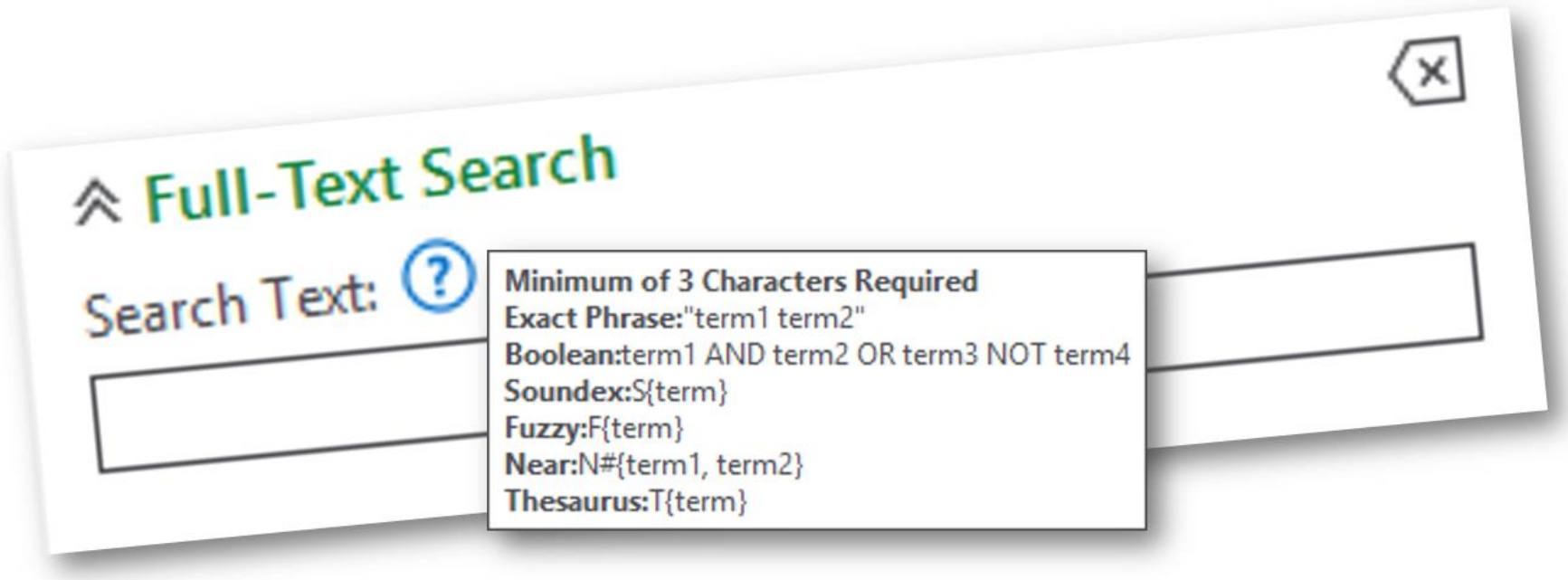
- API più semplici
- Interazione con linguaggi non-Java/JVM
- Facilità operativa di utilizzo
- Clustering e Repliche



Ricerca full-text

La ricerca **full-text** (detta ricerca a testo intero) permette di recuperare facilmente un documento partendo da un termine specifico.

In una ricerca **full-text** il motore di ricerca esamina ogni parola in ciascun documento archiviato con il fine di trovare un riscontro secondo determinati criteri.

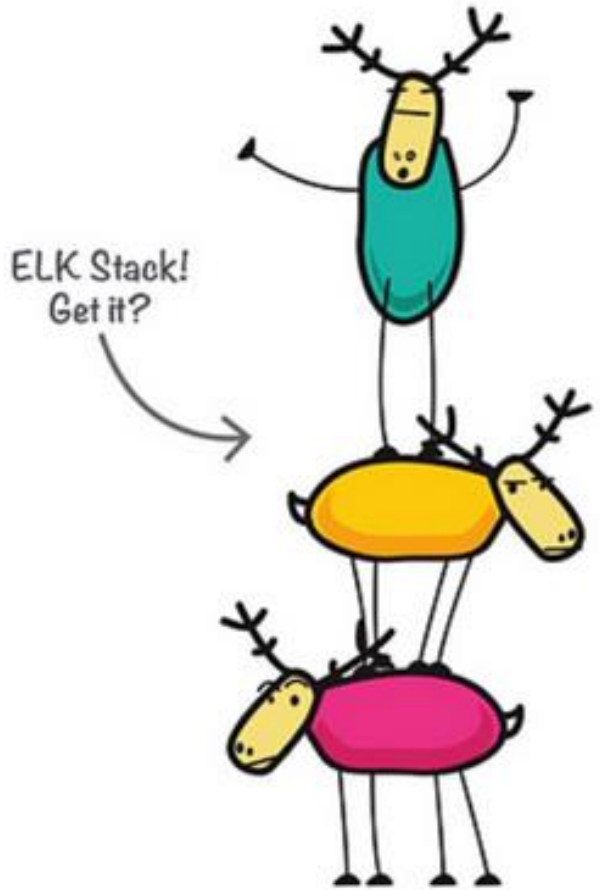


E' in grado di indicizzare rapidamente enormi quantità di dati sia strutturati che non strutturati forniti al sistema sotto forma di isole dati JSon schema-less.

https://www.json.org/json-it.html

```
"bands": [{
    "name": "Metallica",
    "description": "Metallica is an American heavy metal band. The band was formed in 1981 i
        Los Angeles by vocalist/guitarist James Hetfield and drummer Lars Ulrich, and has
        been based in San Francisco for most of its career.",
    "image": "metallica.jpg",
    "year_established": 1981,
    "albums": [{
            "name": "Kill 'Em All",
            "year": 1983
        },
        {
            "name": "Ride the Lightning",
            "year": 1984
        }
    ]
},
{
    "name": "U2",
    "description": "U2 are an Irish rock band from Dublin, formed in 1976. The group consist
        of Bono (lead vocals and rhythm guitar), the Edge (lead guitar, keyboards, and
        backing vocals), Adam Clayton (bass guitar), and Larry Mullen Jr. (drums and
        percussion).",
    "image": "u2.jpg",
    "year_established": 1976,
    "albums": [{
            "name": "Boy",
            "year": 1980
        },
        {
            "name": "October",
            "year": 1981
        }
    ]
}]
```

Quando **Elasticsearch** viene usato unitamente a **Logstash** e **Kibana**, i tre tools formano lo «**Stack ELK**».



E Elasticsearch

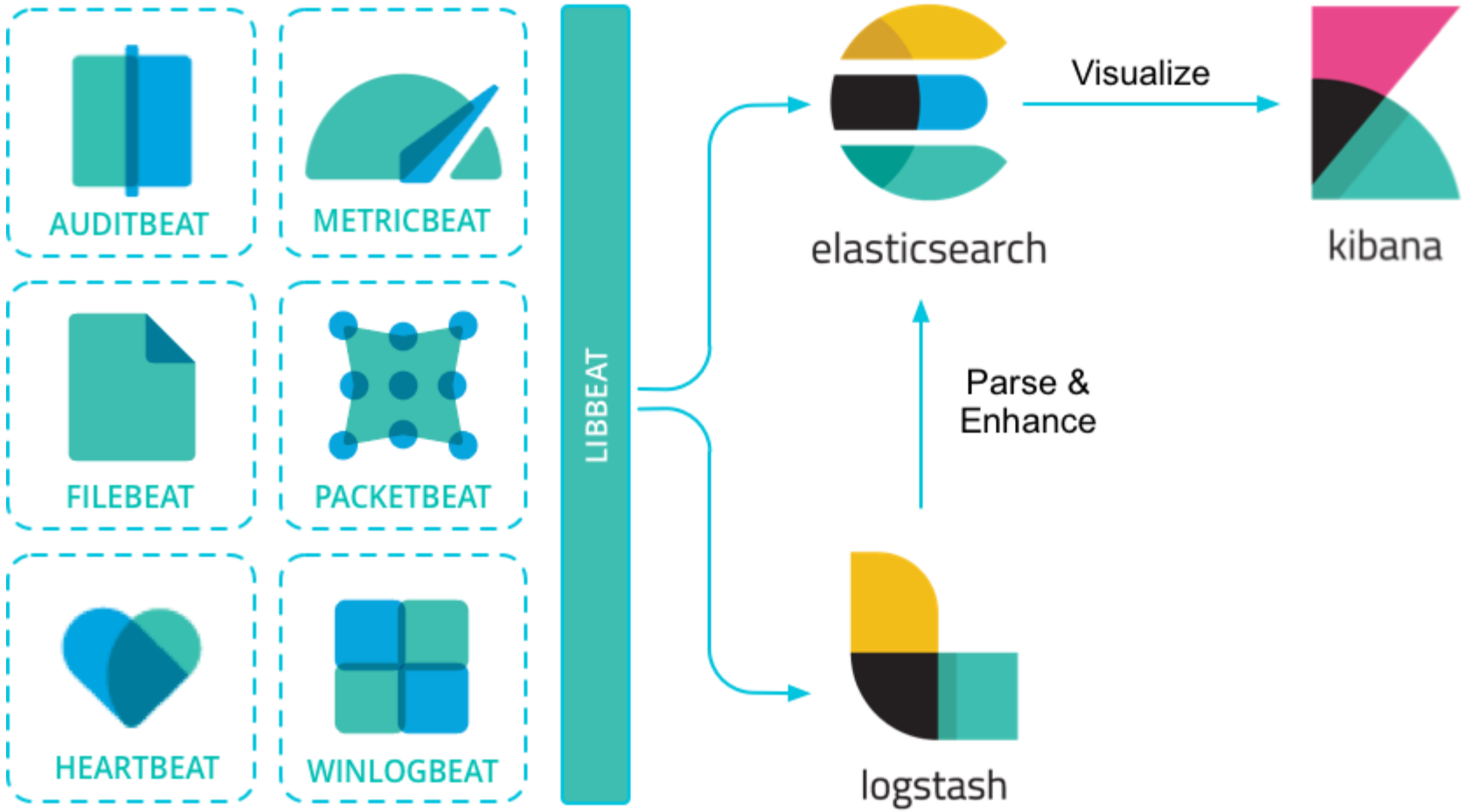
L Logstash

K Kibana

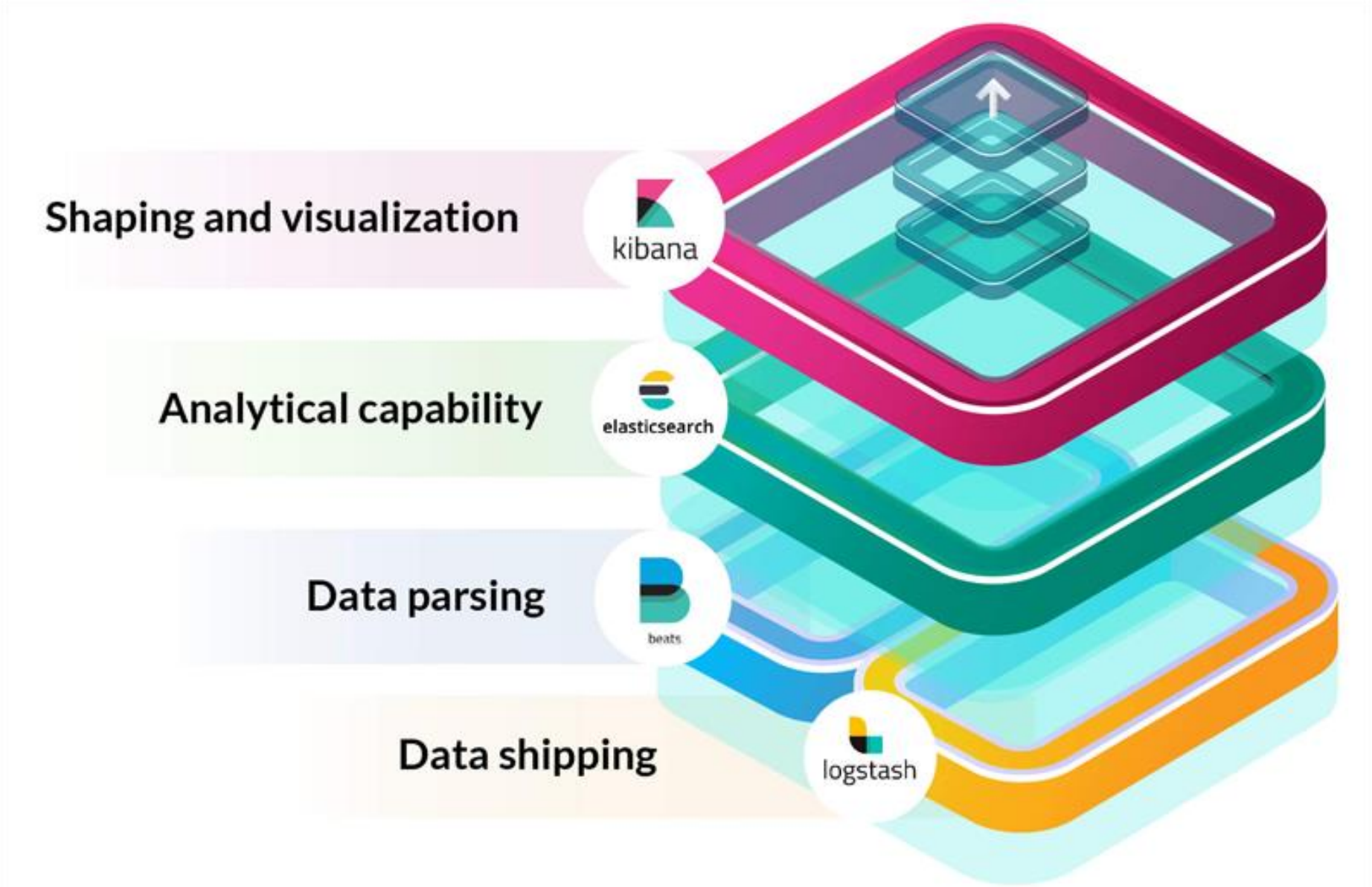
Quando lo **Stack ELK** prevede anche l'uso di Beats,
l'insieme dei quattro tools viene definito più genericamente «**Elastick Stack**».

Data Collection	Data Aggregation & Processing	Indexing & storage	Analysis & visualization

Architettura Elastic Stack



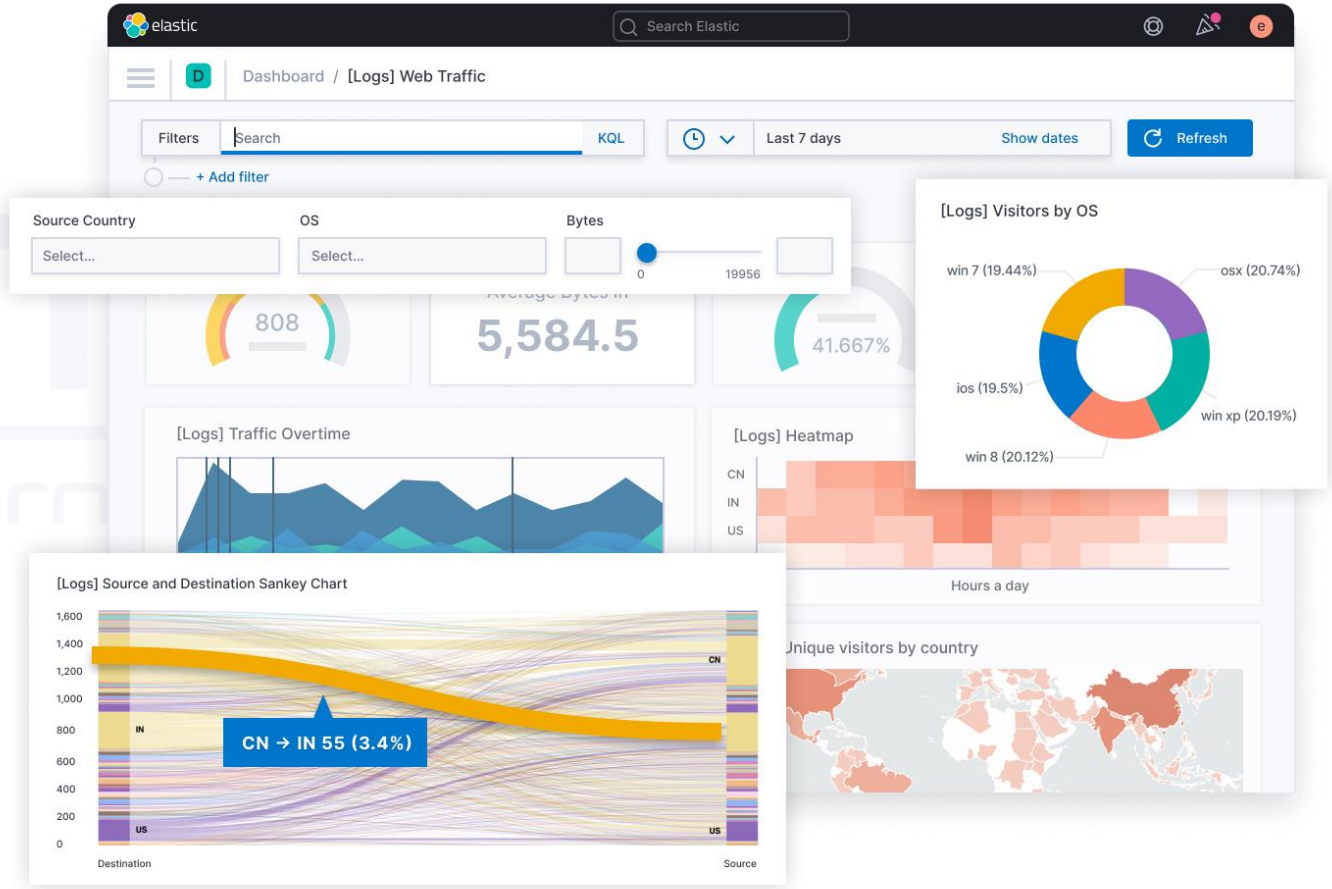
Architettura Elastic Stack

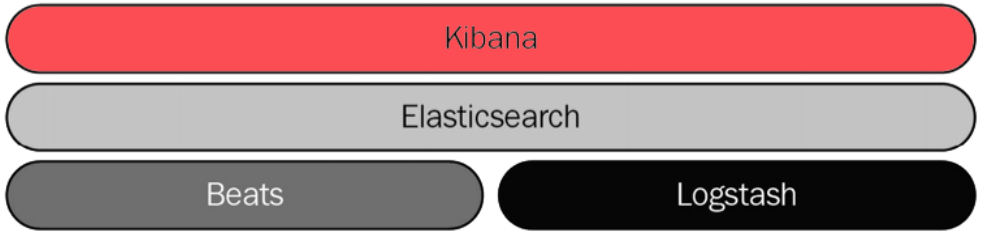


Kibana è uno strumento che permette di esplorare, visualizzare e costruire **dashboard** sui dati salvati in Elasticsearch.

La caratteristica principale di Kibana è l'interrogazione e l'analisi dei **dati**.

Permette di visualizzare i dati in modi alternativi utilizzando mappe di calore, grafici a linee, istogrammi, grafici a torta e mappe geospaziali.

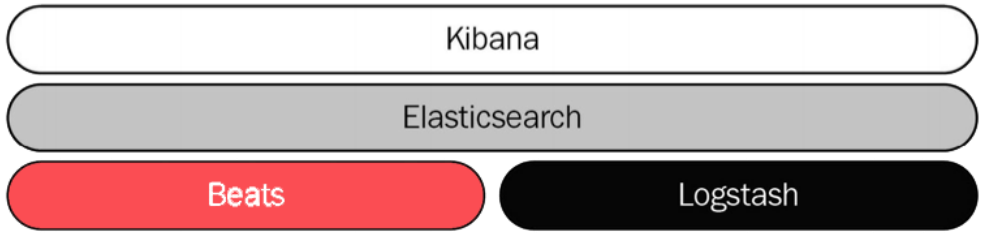


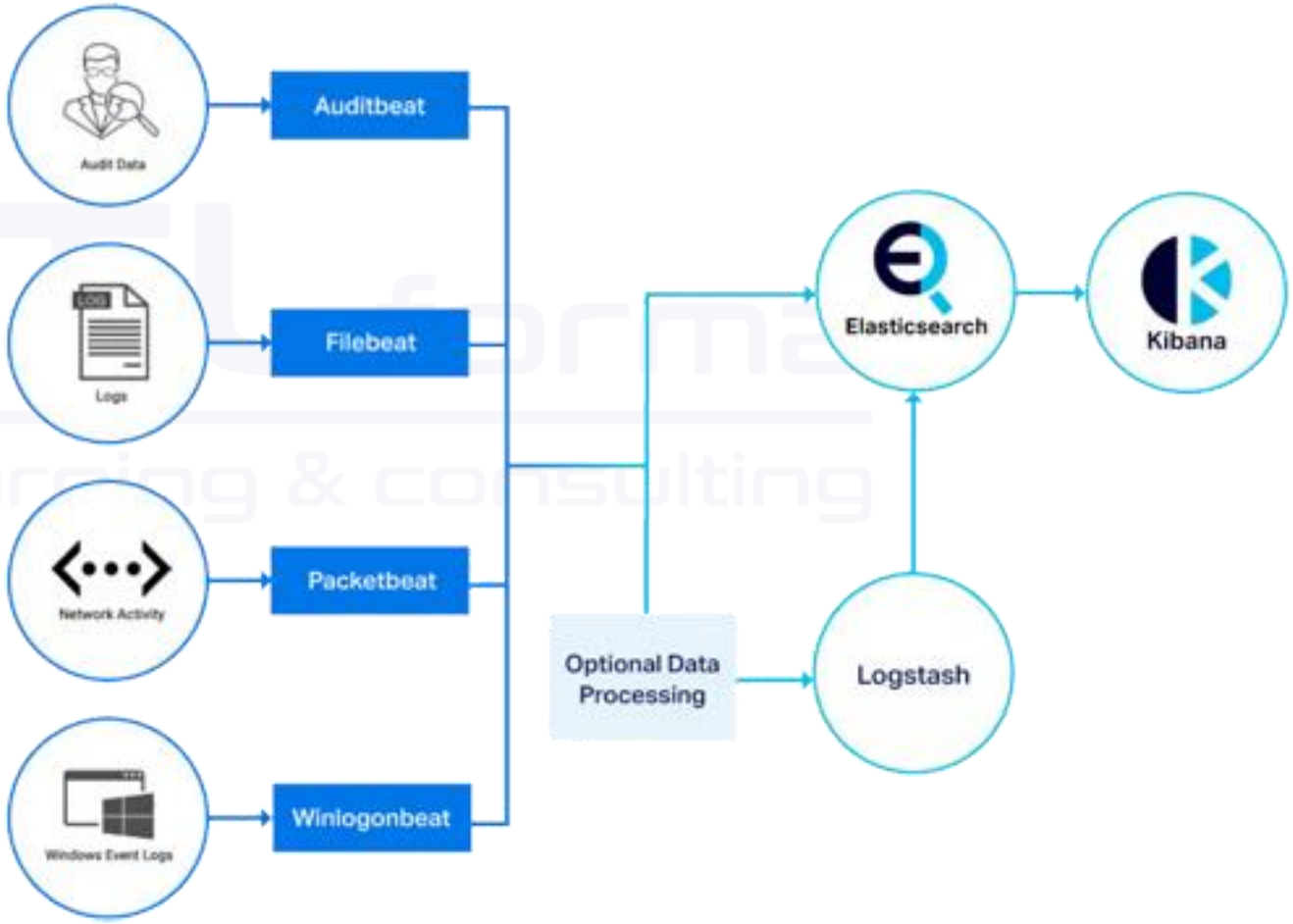


I «beats» sono un sofisticato sistema di shippers («spedizionieri» di dati installati come agents sui sistemi da analizzare) che permettono il collect (raccolta) e l'invio dei dati verso lo stack ELK.

Lista di alcuni dei principali beats agent:

- Auditbeat (audit data)
- Filebat (log files)
- Functionbeat (cloud data)
- Heartbeat (availability)
- Journalbeat (systemd journals)
- Metricbeat (metrics)
- Packetbeat (network traffic)
- Winlogbeat (Windows event logs)

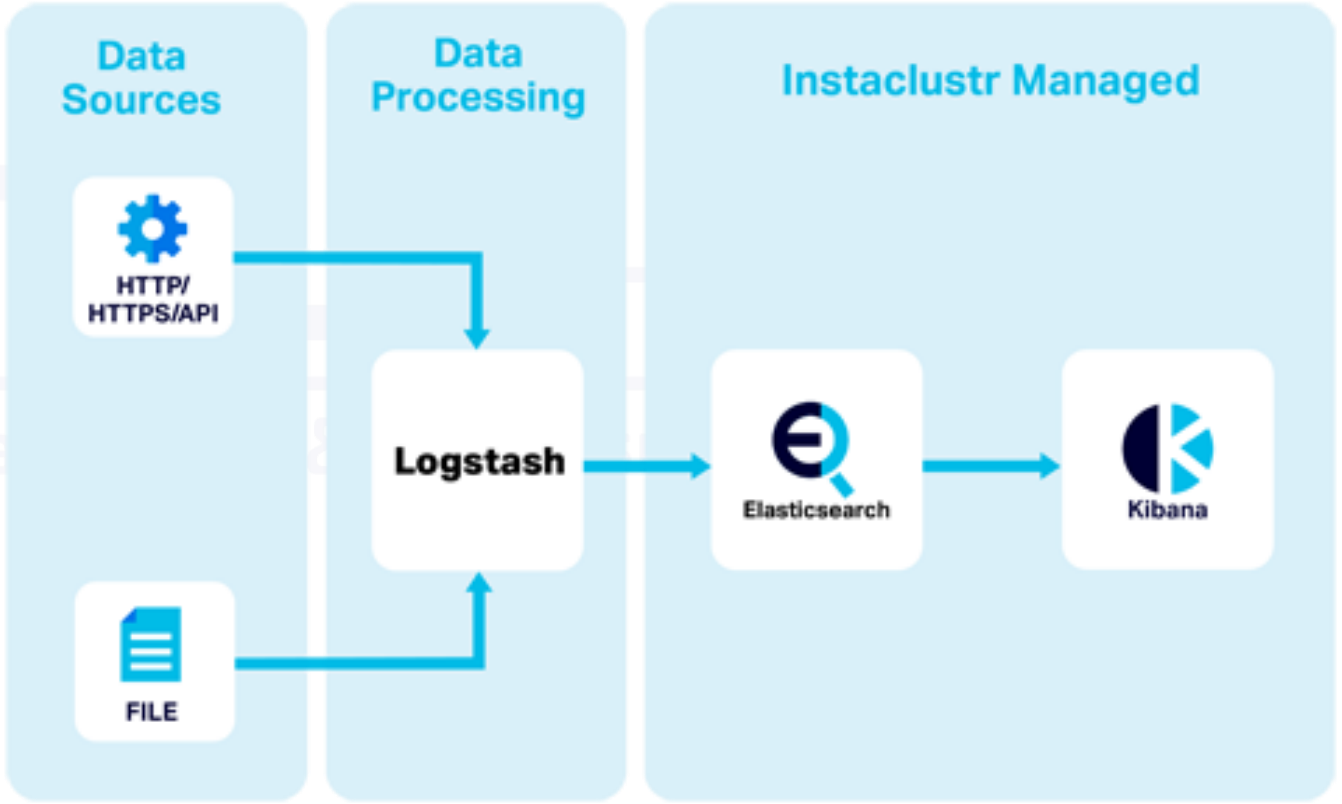


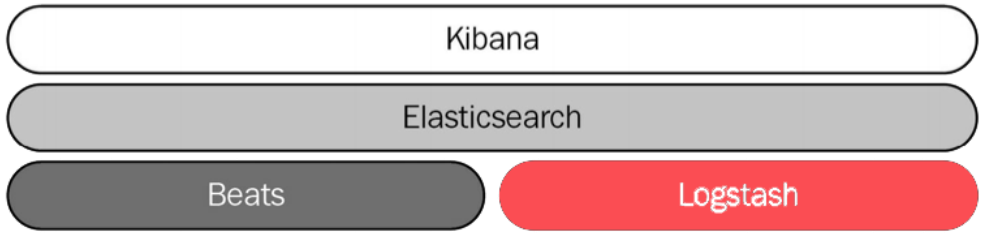


Logstash è uno strumento di ETL (Extract, Transform, Load) utilizzato per elaborare e importare dati da varie fonti (sorgenti).

Le fonti possono essere di varia natura:

- Log Files
- Beats Agent
- Message queues
- Streaming Platforms
- Etc..





Ogni componente risolve un singolo problema comune relativo ai **dati**.

La genericità rende lo stack **flessibile** e **indipendente** dal dominio di applicazione, consentendone l'adozione nei più svariati scenari di utilizzo:

- Ricerca Dati
- SIEM
- Machine Learning
- Analisi delle Performance
- Monitoring
- Alerting
- Analitycs
- Etc.

Evoluzione dell'Elastic Stack

L'ideatore nonchè creatore di Elasticsearch è **Shay Banon**.

Attualmente ricopre la carica di **CTO** nella company **Elasticsearch BV** che egli stesso ha fondato (e di cui è stato anche **CEO**) per promuovere Elasticsearch, relative soluzioni commerciali, e altri software correlati.

Elastichsearch nasce come evoluzione di un altro progetto di Banon denominato «**Compass**» del 2004.

Compass è un framework open source java di tipo Transactional Object/Search Engine Mapping (OSEM) costruito su Lucene.

Nascita di Elasticsearch

elasticsearch

Nel 2019, **Banon** si rese conto della necessità di evolvere Compass per cui decise di **riscriverlo** al fine di aggiungere funzionalità native come:

- Clusterizzazione
- Scalabilità
- Integrazione tramite API RESTful su HTTP
- Etc.

...nacque così Elasticsearch !

Elasticsearch fa parte di quella famiglia di prodotti definiti: «**Distributed Search Engine**».

elasticsearch

- Elasticsearch viene integrato con **Logstash**.
- Viene sviluppato **Kibana** per fornire una UI e per facilitare l'utilizzo delle **API Rest** di elastichsearch. Con Kibana e Logstash nasce l' «**ELK Stack** ».
- Nasce **Elastic Cloud Enterprise** (ECE).
- Viene creato **Packetbeat**, un tool open source usato per raccogliere e inviare «network packet data» a Elasticsearch.
- **Packetbeat** si evolve nel progetto **Beats**, una raccolta di «**agents**» leggeri progettati per raccogliere e spedire diversi tipi di dati verso Elasticsearch.

elasticsearch

- Vengono aggiunte funzionalità di «**Machine Learning**» a Elasticsearch e Kibana per supportare i casi d'uso di rilevamento delle anomalie sui dati che risiedono su Elasticsearch.

- Viene aggiunta l'app «**APM**» (Application Performance Monitoring) su Kibana. APM unita alle app **Logs**, **Metrics** e **Uptime** costituiscono la macro-funzione «**Observability**» di ELK.

- Kibana viene arricchito con funzionalità **SIEM** (Security Information and Event Management) e di analisi della **sicurezza**.

elasticsearch

- Vengono aggiunte una serie di features commerciali e proprietarie sotto forma di una raccolta di componenti denominata «X-Pack».
- Viene aggiunta la funzionalità «Endpoint Detection and Response» che insieme alle funzionalità SIEM vanno a costituire la soluzione ELK per la sicurezza.
- Alla soluzione «Enterprise» vengono aggiunte funzionalità out-of-the-box per la ricerca di siti Web, contenuti etc.