elastic

Elastic Stack
Data Ingestion & Visualizations

Aravind Putrevu
Developer | Evangelist
@aravindputrevu | aravindputrevu.in

Agenda

1. The Ingest Story
2. Logstash: Only for Logs?
3. Configuring and Managing Logstash
4. Visualizing Data Ingested in Kibana
5. Demo

elastic

Agenda

1. The Ingest Story
2. Logstash: Only for Logs?
3. Configuring and Managing Logstash
4. Visualizing Data Ingested in Kibana
5. Demo

elastic

Agenda

1. The Ingest Story
2. Logstash: Only for Logs?
3. Configuring and Managing Logstash
4. Visualizing Data Ingested in Kibana
5. Demo

elastic

Agenda

1. The Ingest Story
2. Logstash: Only for Logs?
3. Configuring and Managing Logstash
4. Visualizing Data Ingested in Kibana
5. Demo

elactic

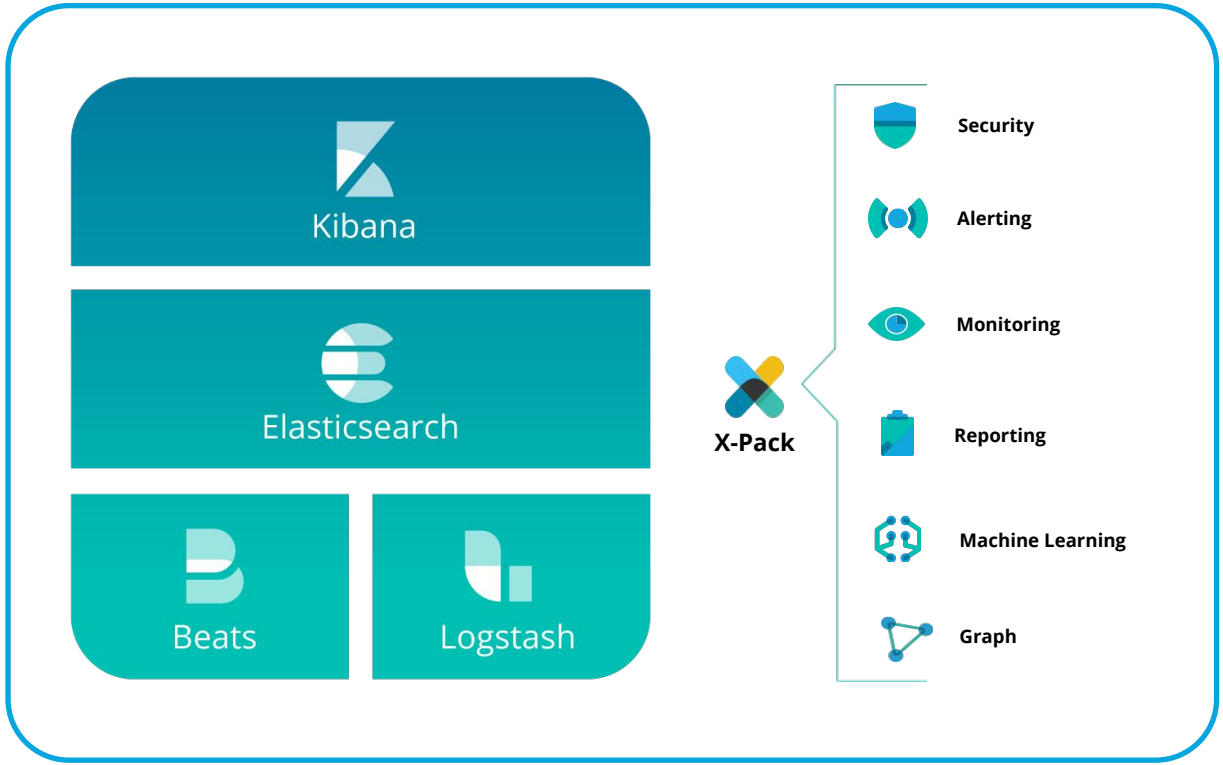
Agenda

1. The Ingest Story
2. Logstash: Only for Logs?
3. Configuring and Managing Logstash
4. Visualizing Data Ingested in Kibana
5. Demo

elastic

Elastic Stack

No enterprise edition
All new versions with 6.2



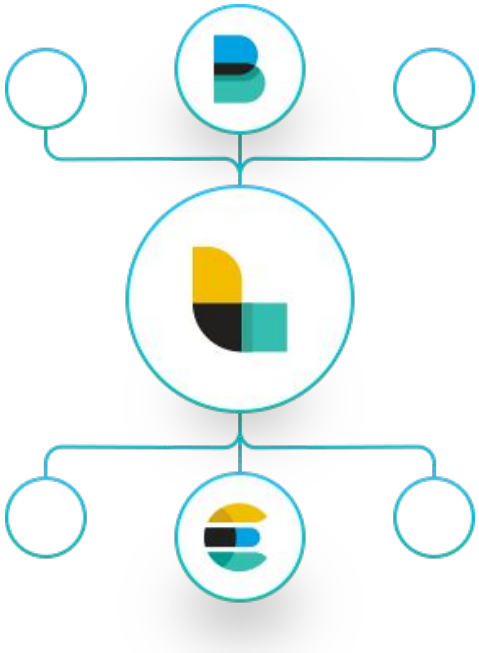
elastic

SECURITY ANALYTICS	LOG ANALYTICS	METRICS ANALYTICS	BUSINESS ANALYTICS	SEARCH	APM
Protect your data	Be alerted on changes	Detect anomalies	Monitor your Elastic Stack	Find links in your data	Share your insights

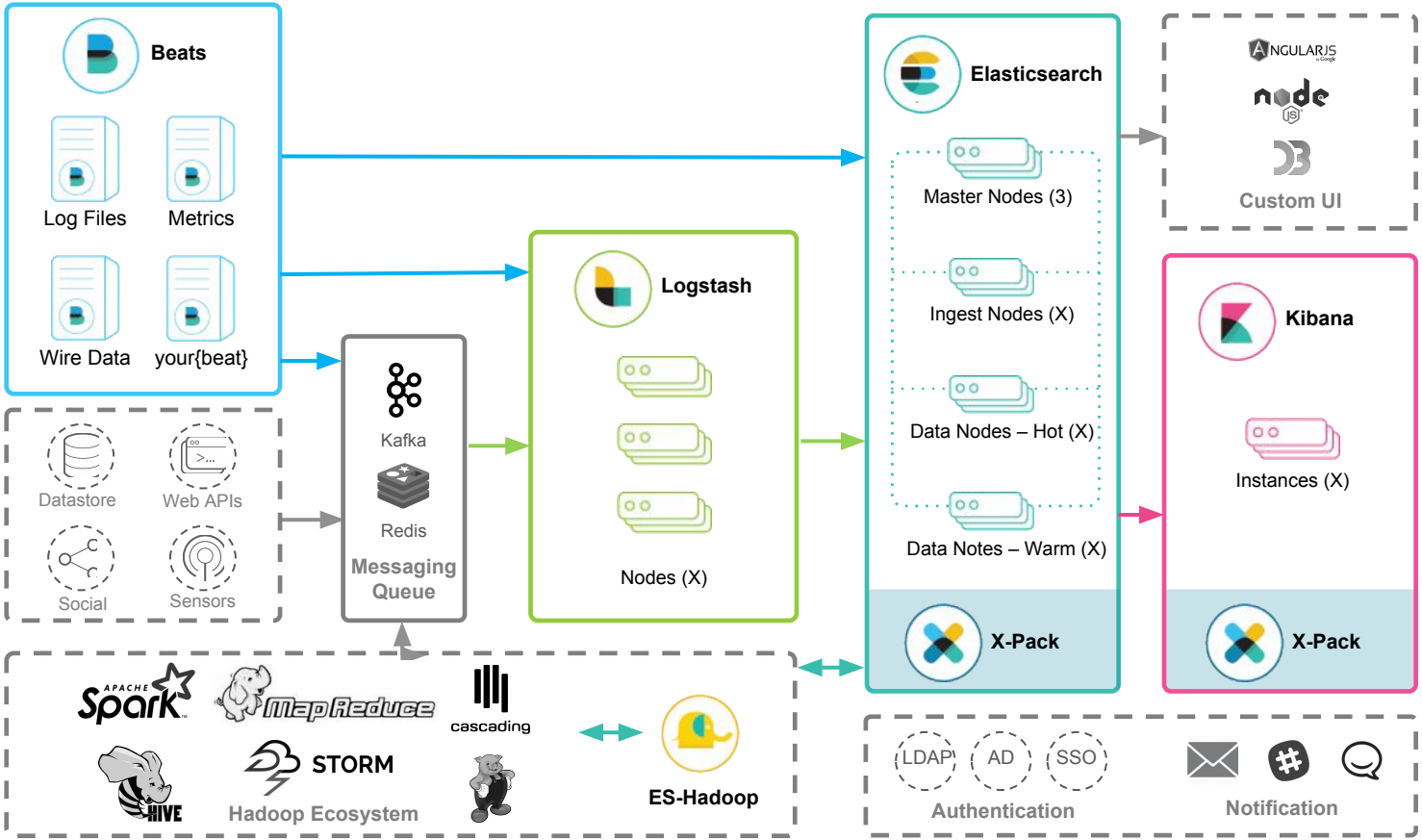
elastic

Logstash

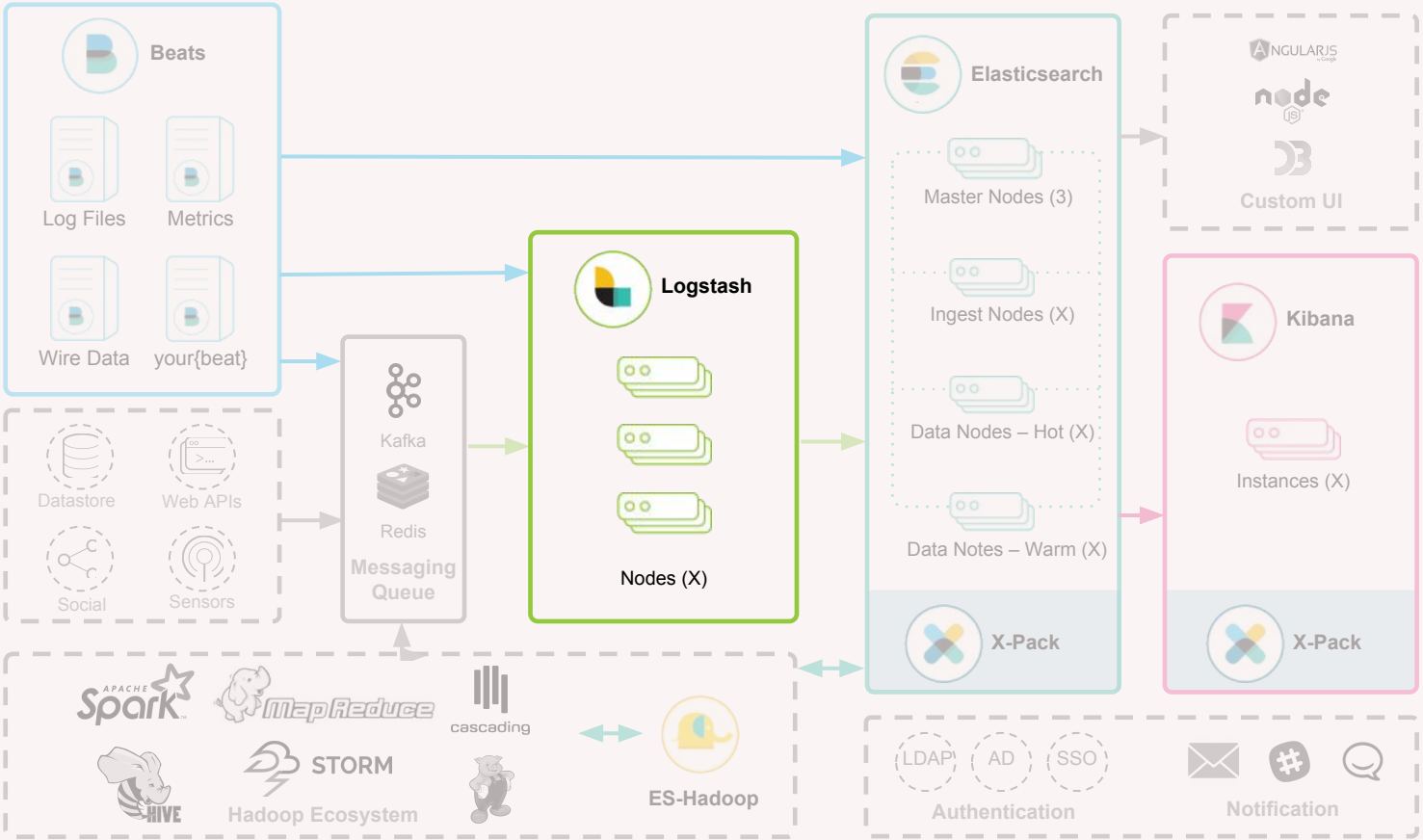
Data processing pipeline



Ingest data of all shapes, sizes, and sources	Parse and dynamically transform data	Transport data to any output
Secure and encrypt data inputs	Build your own pipeline	More than 200+ plugins

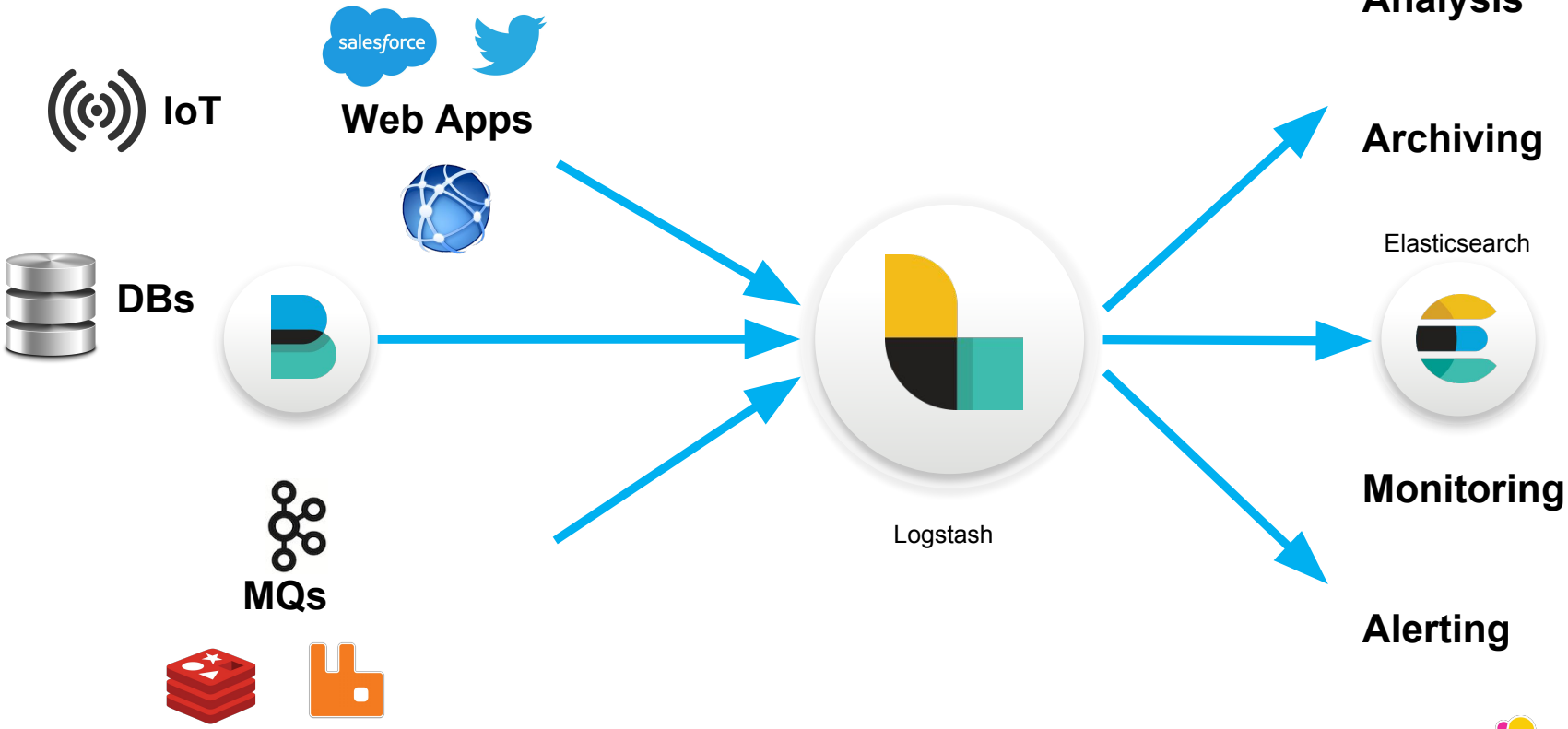

Beats
Log Files
Metrics
Wire Data
your{beat}
Datastore
Web APIs
Social
Sensors
Kafka
Redis
Messaging Queue
Logstash
Nodes (X)
Elasticsearch
Master Nodes (3)
Ingest Nodes (X)
Data Nodes – Hot (X)
Data Notes – Warm (X)
X-Pack
ANGULARJS
node
Custom UI
Kibana
Instances (X)
X-Pack
APACHE Spark
MapReduce
cascading
HIVE
STORM
Hadoop Ecosystem
ES-Hadoop
LDAP
AD
SSO
Authentication
Notification

elastic


Beats
Log Files
Metrics
Wire Data
your{beat}
Datastore
Web APIs
Social
Sensors
Kafka
Redis
Messaging Queue
Logstash
Nodes (X)
Elasticsearch
Master Nodes (3)
Ingest Nodes (X)
Data Nodes – Hot (X)
Data Notes – Warm (X)
X-Pack
ANGULARJS
node
Custom UI
Kibana
Instances (X)
X-Pack
APACHE Spark
MapReduce
cascading
HIVE
STORM
Hadoop Ecosystem
ES-Hadoop
LDAP
AD
SSO
Authentication
Notification

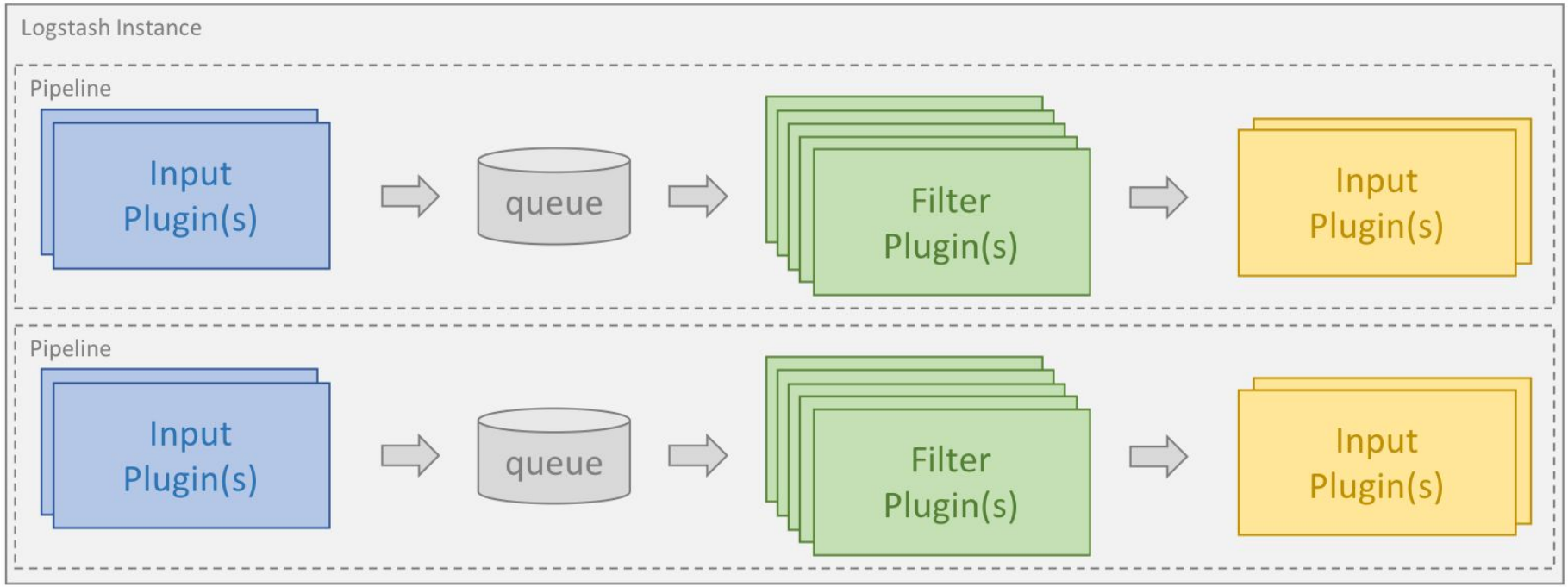
elastic

Popular Data Sources



elastic

How it works?


Logstash Instance
Pipeline
Input Plugin(s)
queue
Filter Plugin(s)
Input Plugin(s)
Pipeline
Input Plugin(s)
queue
Filter Plugin(s)
Input Plugin(s)

Configuring Logstash

```
# This is a comment. You should use comments to describe
# parts of your configuration.
input {
  ...
}

filter {
  ...
}

output {
  ...
}
```

Configuring Logstash

client ip

timestamp

127.0.0.1 - - [11/Dec/2013:00:01:45 -0800] "GET /xampp/status.php HTTP/1.1" 200 3891

"http://cadenza/xampp/navi.php" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10.9; rv:25.0) Gecko/20100101 Firefox/25.0"

request

agent

Configuring Logstash

```
input { stdin { } }

filter {
  grok {
    match => { "message" => "%{COMBINEDAPACHELOG}" }
  }
  date {
    match => [ "timestamp" , "dd/MMM/yyyy:HH:mm:ss Z" ]
  }
}

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

Configuring Logstash

```
{
        "message" => "127.0.0.1 - - [11/Dec/2013:00:01:45 -0800] \"GI
     "@timestamp" => "2013-12-11T08:01:45.000Z",
       "@version" => "1",
           "host" => "cadenza",
       "clientip" => "127.0.0.1",
          "ident" => "-",
           "auth" => "-",
      "timestamp" => "11/Dec/2013:00:01:45 -0800",
           "verb" => "GET",
        "request" => "/xampp/status.php",
    "httpversion" => "1.1",
       "response" => "200",
          "bytes" => "3891",
       "referrer" => "\"http://cadenza/xampp/navi.php\"",
          "agent" => "\"Mozilla/5.0 (Macintosh; Intel Mac OS X 10.9;
}
```

Grok

```
2016-09-19T18:19:00 [8.8.8.8:prd] DEBUG this is an example log message
```

```
%{TIMESTAMP_ISO8601:timestamp} \[%{IPV4:ip};%{WORD:environment}\] %{LOGLEVEL:log_level} %{GREEDYDATA:mess
age}
```

```
{
  "timestamp": "2016-09-19T18:19:00",
  "ip": "8.8.8.8",
  "environment": "prd",
  "log_level": "DEBUG",
  "message": "this is an example log message"
}
```


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

Best Practices for Grok

- Grok may not perform well when a match fails
- Monitor the occurrence of _grokparsefailures and then benchmark their cost
- Use anchors such as ^ and $ to remove ambiguity and aid the regex engine
- Tiered matching increases performance if you don't use anchors, otherwise don't bother. When in doubt, measure!
- Use Monitoring or Metrics API.

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

Grok Debugger

Dev Tools

Console | Search Profiler | Grok Debugger

Sample Data

```
55.3.244.1 DEBUG This is Grok Test!
```

Grok Pattern

```
%{IP:client} %{WORD:loglevel} %{GREEDYDATA:message}
```

Custom Patterns

Simulate

Structured Data

```
{
  "loglevel": "DEBUG",
  "client": "55.3.244.1",
  "message": "This is Grok Test! "
}
```


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

elastic

Pipelines in 1, 2, 3

Start Fast, Think Big

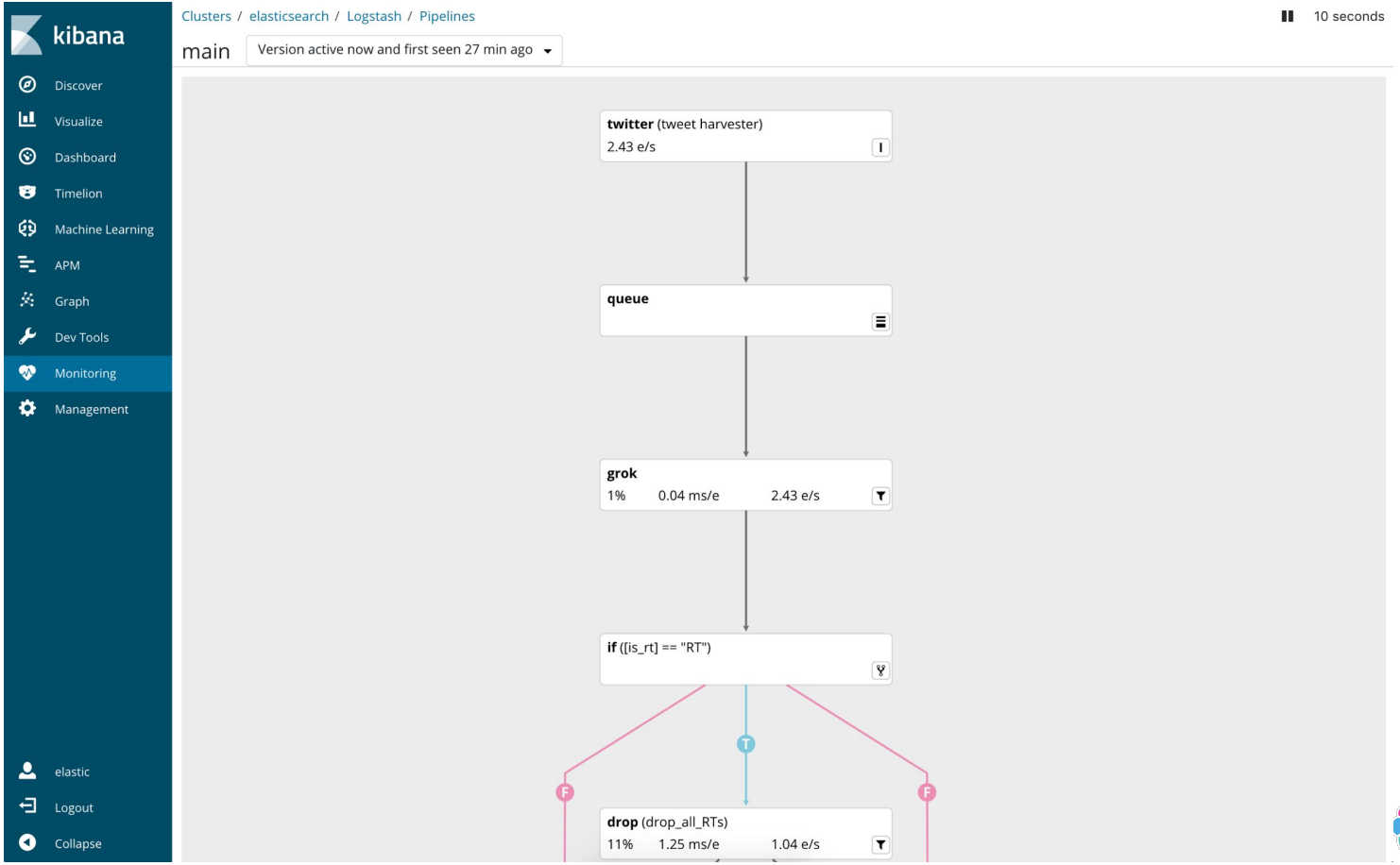
- Configure your pipelines, not code them
- Stash your first event in minutes
- Grow iteratively, scale horizontally

elastic

Multiple Pipelines

```
- pipeline.id: my-pipeline_1
  path.config: "/etc/path/to/p1.config"
  pipeline.workers: 3
- pipeline.id: my-other-pipeline
  path.config: "/etc/different/path/p2.cfg"
  queue.type: persisted
```

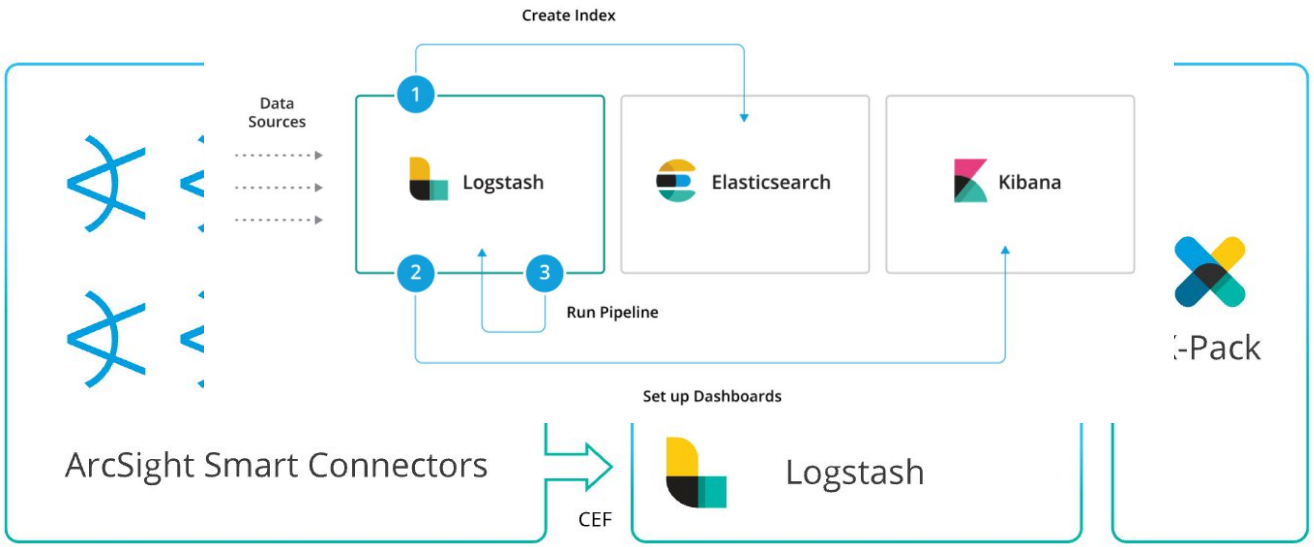
Centralized Pipeline Management



elastic

Logstash Modules

- Elastic Cloud
- Arcsight
- Netflow



Dashboard / Netflow: Geo Location

Share Clone Edit Today

*

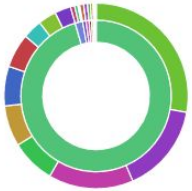
Uses lucene query syntax

Add a filter +

Netflow: Dashboard Navigation

Overview | Conversation Partners | Traffic Analysis | Top-N | Geo Location | Autonomous Systems | Flow Exporters | Raw Flow Records

Netflow: Countries and Cities (flow records)



- United States
- Canada
- Australia
- Netherlands
- China
- Ireland
- Antigua and Barbu...
- New Zealand
- Singapore

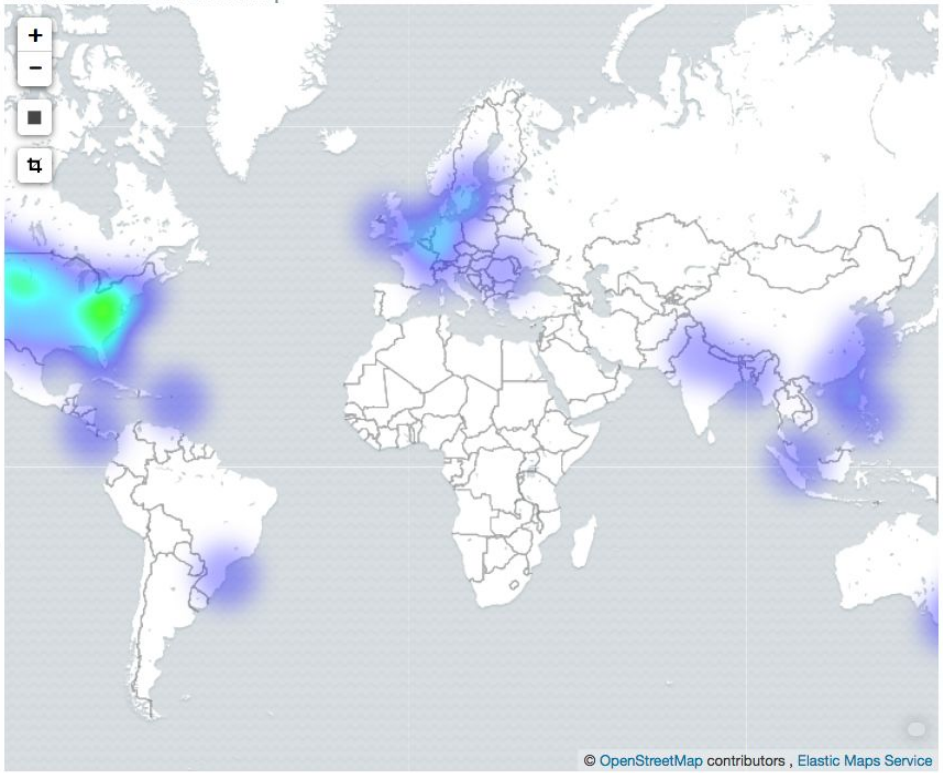
Netflow: Destinations and Sources (flow records)

- 172.16.139.250
- 172.16.133.6
- 8.8.8.8
- 172.16.128.169
- 172.16.133.54
- 172.16.128.202
- 172.16.133.132
- 172.16.133.41
- 172.16.133.254

Netflow: Destination and Source Ports (flow rec...

- TCP/5440
- TCP/80 (http)
- TCP/443 (https)
- UDP/53 (dns)
- UDP/161 (snmp)
- UDP/1853 (vids-avtp)
- UDP/1900 (ssdp)
- UDP/8200 (trivnet1)
- UDP/137 (netbios-...

Netflow: Geo Location Heatmap

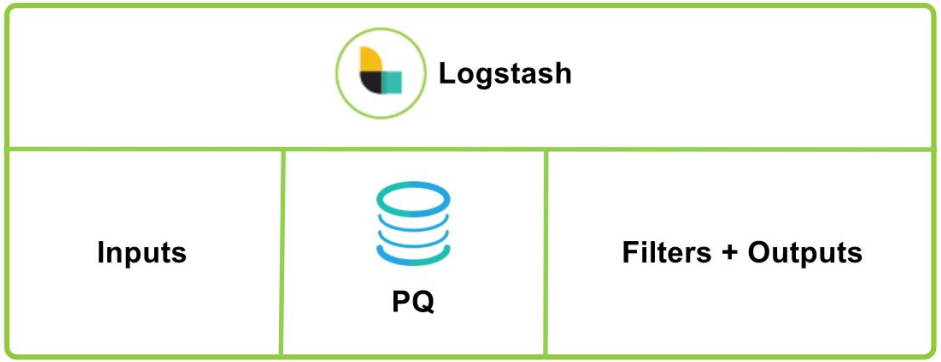


Data Resiliency

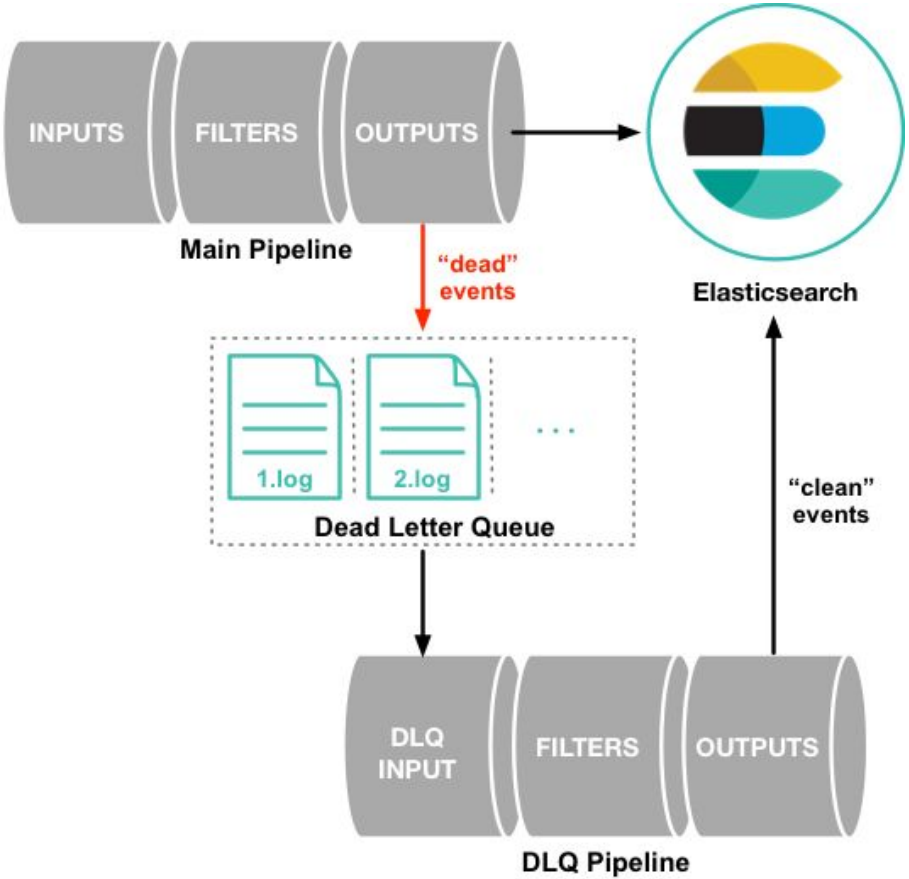
- Persistent Queues
- Dead Letter Queues

Persistent Queue

- Handle bursts of events
- At-least once delivery



Dead Letter Queue



Scaling Logstash

kibana

Discover
Visualize
Dashboard
Timelion
Machine Learning
APM
Graph
Dev Tools
Monitoring
Management

elastic
Logout
Collapse

Elasticsearch cluster nodes have changed: Node was restarted [1]. [0YXPfEc].
Last checked February 5, 2018 1:25:43 PM (triggered 1 min ago)

Elasticsearch • Health is yellow

Overview	
Version	6.2.0
Uptime	2 minutes
Jobs	0

Nodes: 1	
Disk Available	17.65% 41GB / 234GB
JVM Heap	46.22% 458MB / 990MB

Indices: 13	
Documents	26,025
Disk Usage	26MB
Primary Shards	17
Replica Shards	0

Kibana • Health is green

Overview	
Requests	60
Max. Response Time	250 ms

Instances: 1	
Connections	24
Memory Usage	13.44% 192MB / 1GB

Logstash

Overview	
Events Received	180
Events Emitted	180

Nodes: 1	
Uptime	2 minutes
JVM Heap	28.88% 286MB / 990MB

Pipelines: 1

Clusters / elasticsearch / Logstash

10 seconds Last 15 minutes

Overview Nodes Pipelines

Nodes: **1** Memory: **226MB / 990MB** Events Received: **609** Events Emitted: **609**

Events Received Rate (/s)

Events Received Rate 0 /s

Events Emitted Rate (/s)

Events Emitted Rate 0 /s

Event Latency (ms)

Event Latency 0 ms

Overview Pipelines Advanced

127.0.0.1:9600 Events Received: **606** Events Emitted: **606** Config Reloads: **0** Pipeline Workers: **8** Batch Size: **125** Version: **6.0.0** Uptime: **6 minutes**

Events Received Rate (/s)

Events Received Rate 0 /s

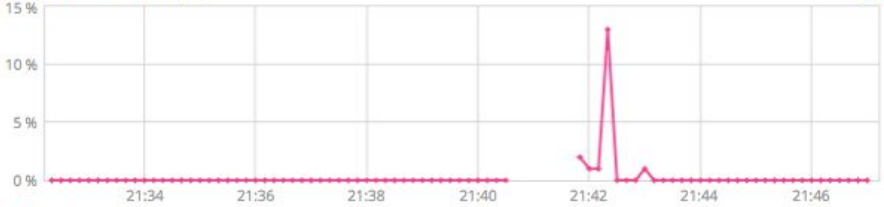
JVM Heap (MB)

Max Heap 989.9 MB Used Heap 342.0 MB

Events Emitted Rate (/s)

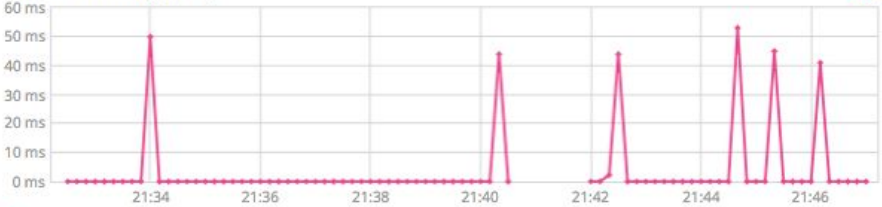
Events Emitted Rate 0 /s

CPU Utilization (%)



CPU Utilization 0 %

Event Latency (ms)



Event Latency 0 ms

System Load

1m 3.59 5m N/A 15m N/A

Measure...Measure...Measure...

kibana

Discover
Visualize
Dashboard
Timelion
Machine Learning
APM
Graph

Clusters / elasticsearch / Logstash

10 seconds Last 30 minutes

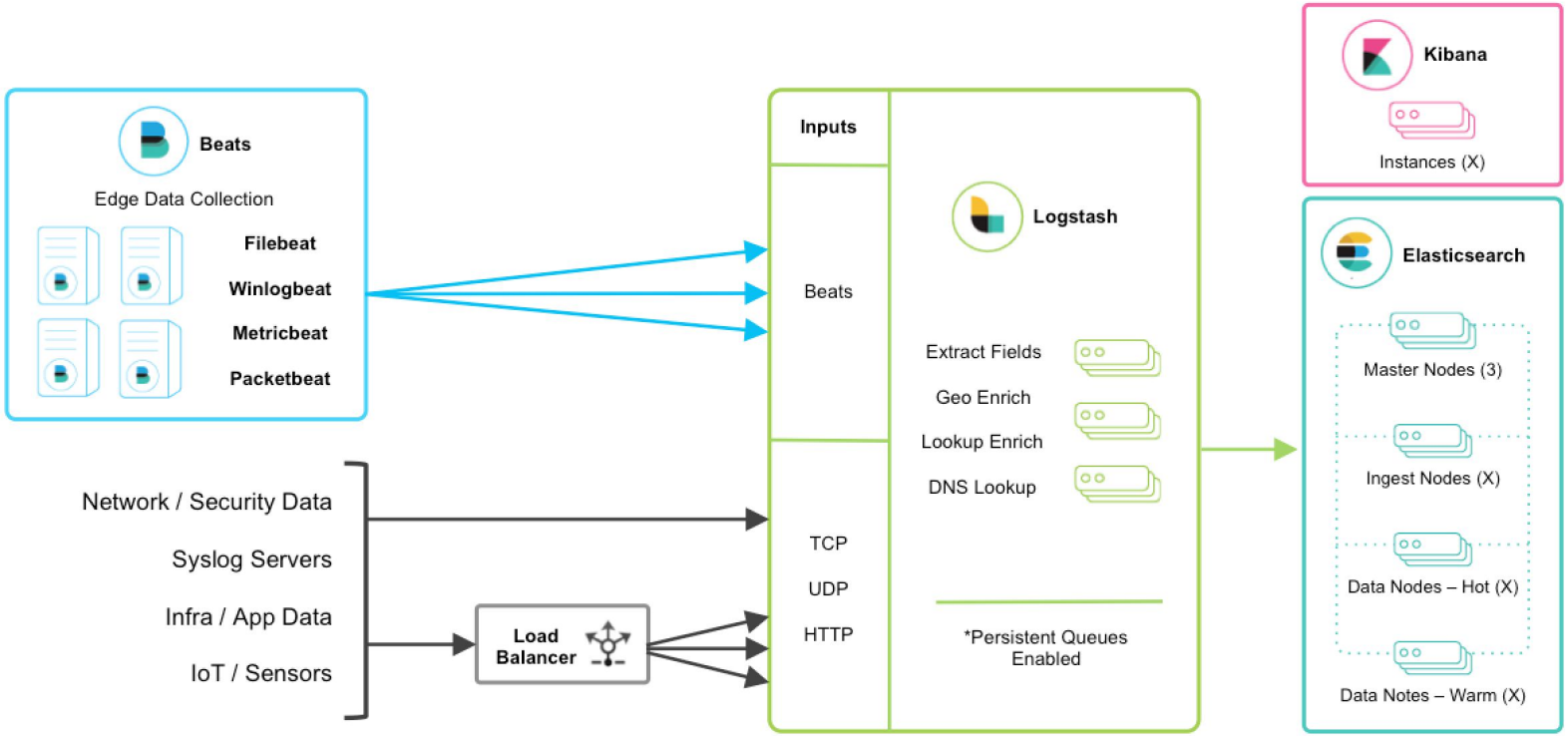
Overview Nodes Pipelines

Nodes: **1** Memory: **321MB / 990MB** Events Received: **3.8k** Events Emitted: **3.8k**

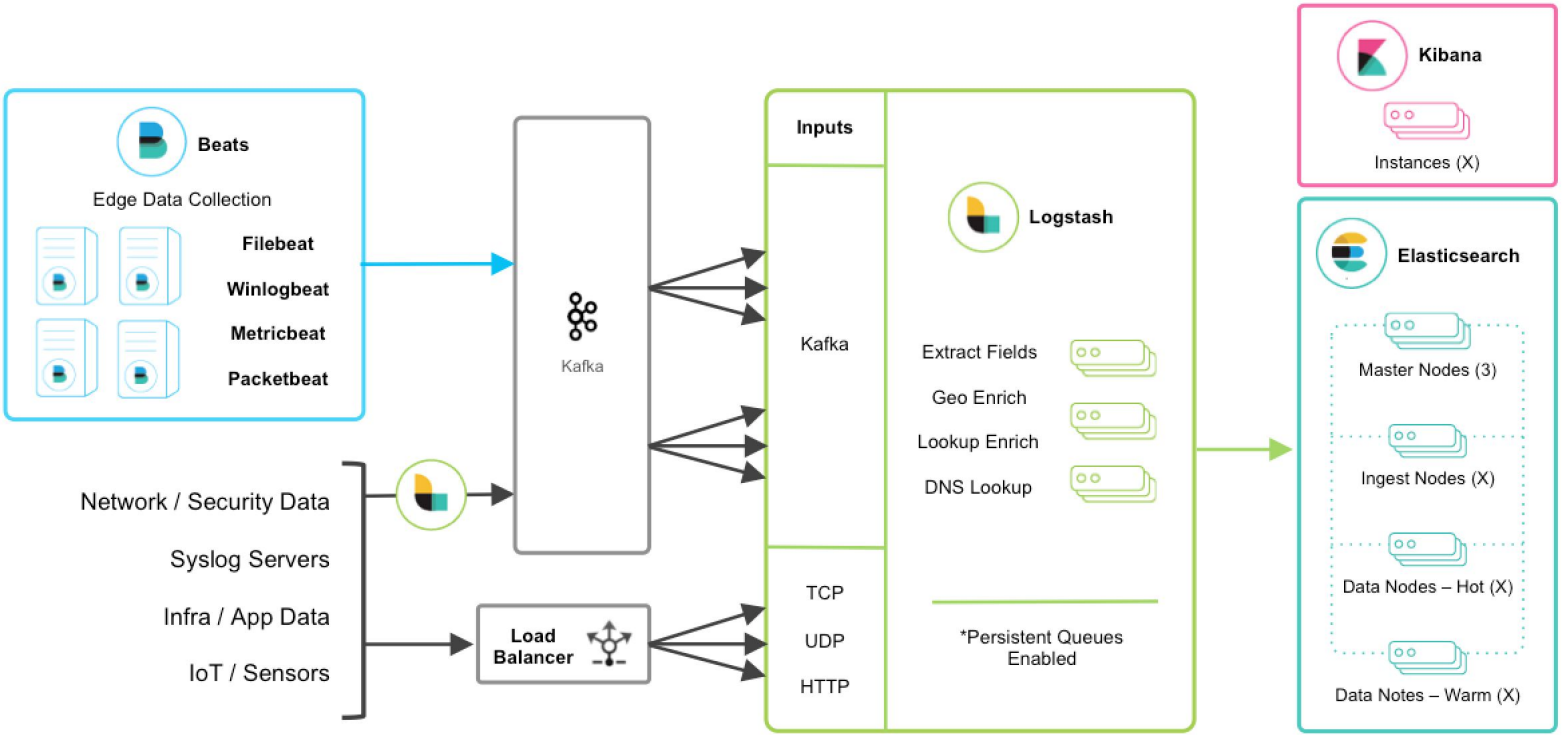
Filter Pipelines...

ID	Events Emitted Rate	Number of Nodes
main	1.2 e/s	1

Scaling Logstash



Scaling Logstash



Logstash Plugins

- Input
- Output
- Filter

https://www.elastic.co/guide/en/logstash-versioned-plugins/current/index.html

Parse/Extract

elastic

The Grok Filter

The go-to-tool for parsing fields

```
filter {
  grok {
    match => {“message” => “%{TIMESTAMP_8601:ts}%{SPACE}%{GREEDYDATA:message}”}
  }
}
```

https://www.elastic.co/guide/en/logstash-versioned-plugins/current/index.html

elastic

The Date Filter

Use data strings to set @timestamp

```
filter {
  date {
    match => ["timestamp_string", "ISO8601"]
  }
}
```

https://www.elastic.co/guide/en/logstash-versioned-plugins/current/index.html

The KV Filter

The easy way to parse data in key/value pairs

```
filter {
  kv {
    source => “message”
    target => “parsed”
    value_split => “:”
  }
}
```

https://www.elastic.co/guide/en/logstash-versioned-plugins/current/index.html

elastic

Core operations

elastic

The Mutate Filter

The go-to-tool for field manipulation in Logstash

- Convert field types (from strings to integers etc.)
- Add/rename/replace/copy fields
- Upper/lowercase transformation
- Join arrays together (useful for Array=>String operations)
- Merge hashes
- Split fields into Arrays
- Strip whitespace

https://www.elastic.co/guide/en/logstash-versioned-plugins/current/index.html

elastic

Core Transformation Filters

- Mutate - Modify / Add Individual Fields
- Split - Divide a single event into multiple events
- Drop - Delete an event

https://www.elastic.co/guide/en/logstash-versioned-plugins/current/index.html

Core Operations Example

```
filter {
  mutate { lowercase => “account” }
  if [type] == “batch” {
    split { field => actions target => action }
  }
  if { “action” =~ /special/ } {
    drop {}
  }
}
```

https://www.elastic.co/guide/en/logstash-versioned-plugins/current/index.html

Search Supportive

elastic

GeoIP Filter

Enrich IP address information

```
filter {
  geoip {
    fields => “my_geoip_field”
  }
}
```

https://www.elastic.co/guide/en/logstash-versioned-plugins/current/index.html

User Agent Filter

Enrich browser user agent information

```
filter {
  useragent {
    source => “useragent”
  }
}
```

https://www.elastic.co/guide/en/logstash-versioned-plugins/current/index.html

Translate Filter

Use local data to map / enrich events

```
filter {
  translate {
    dictionary => [ "100", "Continue",
                    "101", "Switching Protocols",
                    "merci", "thank you",
                    "old version", "new version" ]
  }
}
```

https://www.elastic.co/guide/en/logstash-versioned-plugins/current/index.html

Elasticsearch Filter

Use Elasticsearch Data to Enrich Events

```
elasticsearch {
  hosts => ["es-server"]
  query => "type:start AND operation:%{[opid]}"
  fields => { "@timestamp" => "started" }
}
```

https://www.elastic.co/guide/en/logstash-versioned-plugins/current/index.html

JDBC Streaming Filter

Use a database to enrich events

```
filter {
  jdbc_streaming {
    jdbc_driver_library => "/path/to/mysql-connector-java-5.1.34-bin.jar"
    jdbc_driver_class => "com.mysql.jdbc.Driver"
    jdbc_connection_string => ""jdbc:mysql://localhost:3306/mydatabase"
    jdbc_user => "me"
    jdbc_password => "secret"
    statement => "select * from WORLD.COUNTRY WHERE Code = :code"
    parameters => { "code" => "country_code"}
    target => "country_details"
  }
}
```

https://www.elastic.co/guide/en/logstash-versioned-plugins/current/index.html

elastic

Add Data to Kibana

Use these solutions to quickly turn your data into pre-built dashboards and monitoring systems.

APM

APM automatically collects in-depth performance metrics and errors from inside your applications.

Add APM

Logging

Ingest logs from popular data sources and easily visualize in preconfigured dashboards.

Add log data

Metrics

Collect metrics from the operating system and services running on your servers.

Add metric data

Security Analytics

Centralize security events for interactive investigation in ready-to-go visualizations.

Add security events

Data already in Elasticsearch? Set up index patterns

Visualize and Explore Data

APM
Automatically collect in-depth performance metrics and errors from inside your applications.

Dashboard
Display and share a collection of visualizations and saved searches.

Discover
Interactively explore your data by querying and filtering raw documents.

Graph
Surface and analyze relevant relationships in your Elasticsearch data.

Machine Learning
Automatically model the normal behavior of your time series data to detect anomalies.

Timelion
Use an expression language to analyze time series data and visualize the results.

Visualize

Manage and Administer the Elastic Stack

Console
Skip cURL and use this JSON interface to work with your data directly.

Index Patterns
Manage the index patterns that help retrieve your data from Elasticsearch.

Monitoring
Track the real-time health and performance of your Elastic Stack.

Saved Objects
Import, export, and manage your saved searches, visualizations, and dashboards.

Security Settings
Protect your data and easily manage who has access to what with users and roles.

Watcher
Detect changes in your data by creating, managing, and monitoring alerts.

Resources

- https://www.elastic.co/learn
- https://www.elastic.co/blog/category/engineering
- https://discuss.elastic.co/
- https://fb.com/groups/ElasticIndiaUserGroup
- https://elastic.co/community

Fin!

discuss.elastic.co | aravind@elastic.co | @aravindputrevu