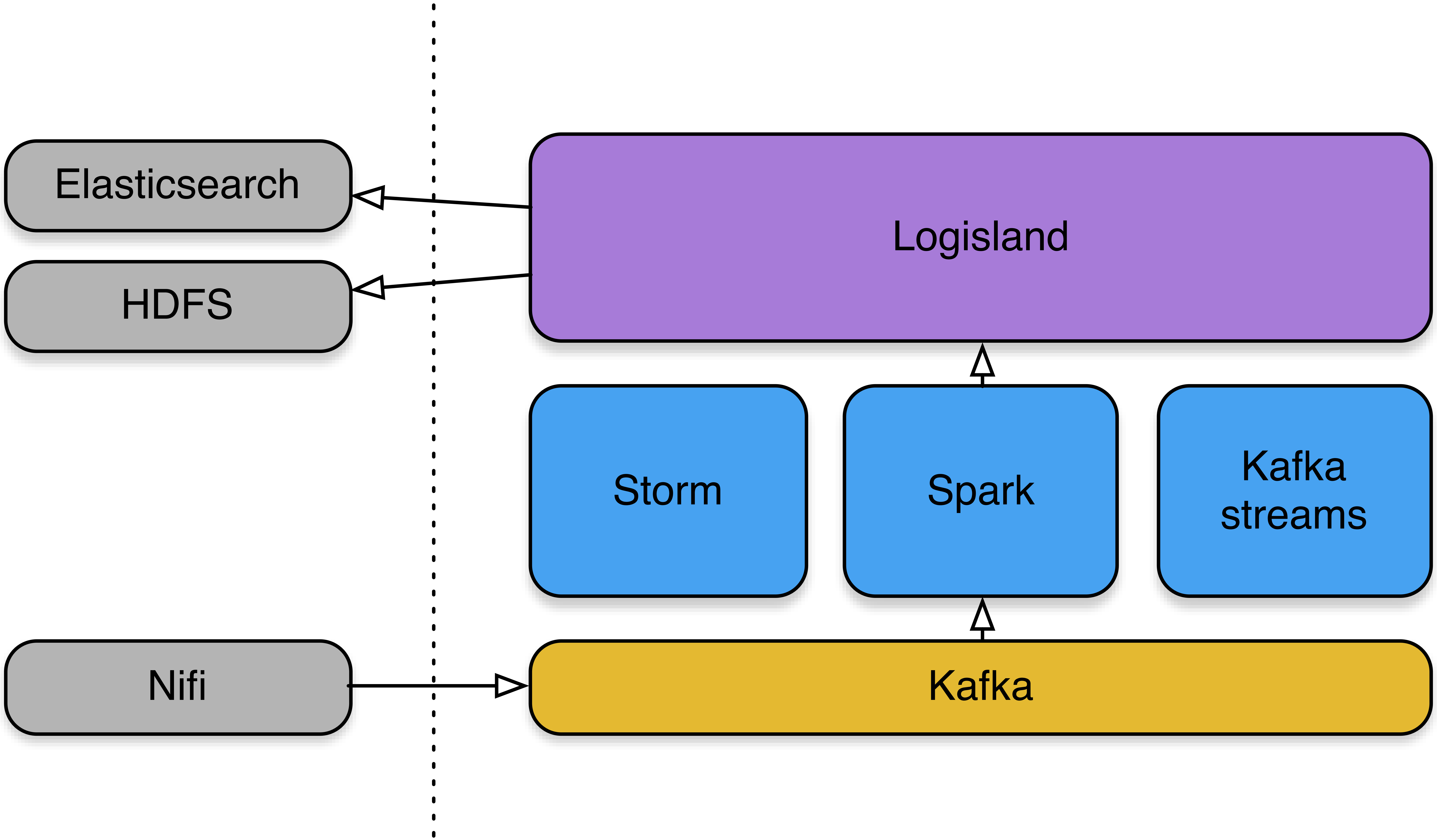
Logisland

Event Pattern Mining

Key features

- Natively **distributed OpenSource framework**.
- Deploy **High level analytics** pipelines with no code
- Use POJO API to easily add **Event mining Processors**
- Compliant with common **Hadoop distrib** (HDP, Cloudera, …)
- Processing framework agnostic (Spark, Storm, Kafka Streams …)
- **Stream taxonomy** with Avro based schemas

Architecture



© Hurence 2014

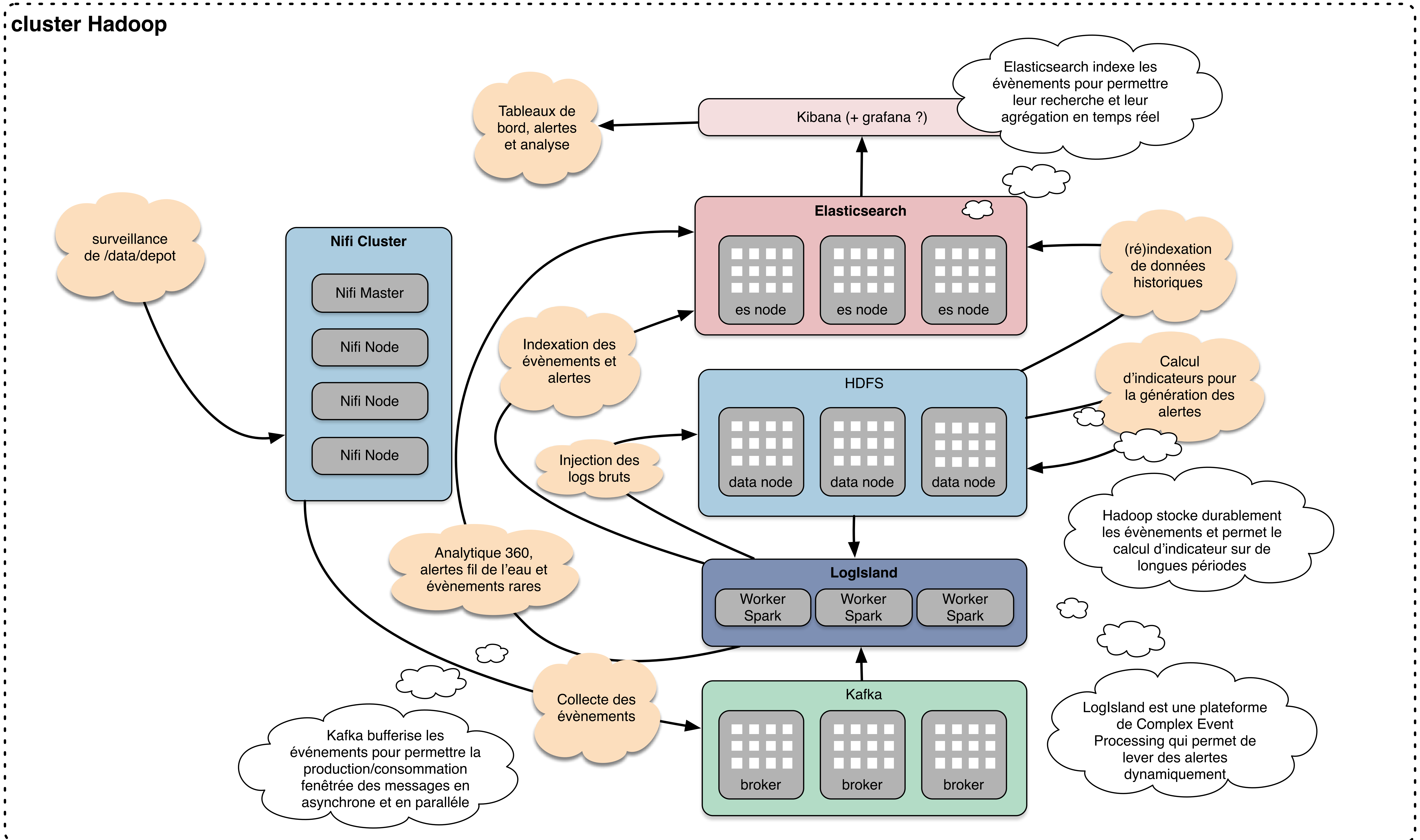
Low level processing

- Multiline regex-based log parsing (applicative logs)
- Index events to search engine (ES / SolR) for low latency graphic analysis
- Store events to distributed filesystem (HDFS) for offline queries and model training
- automatic Timestamp detection

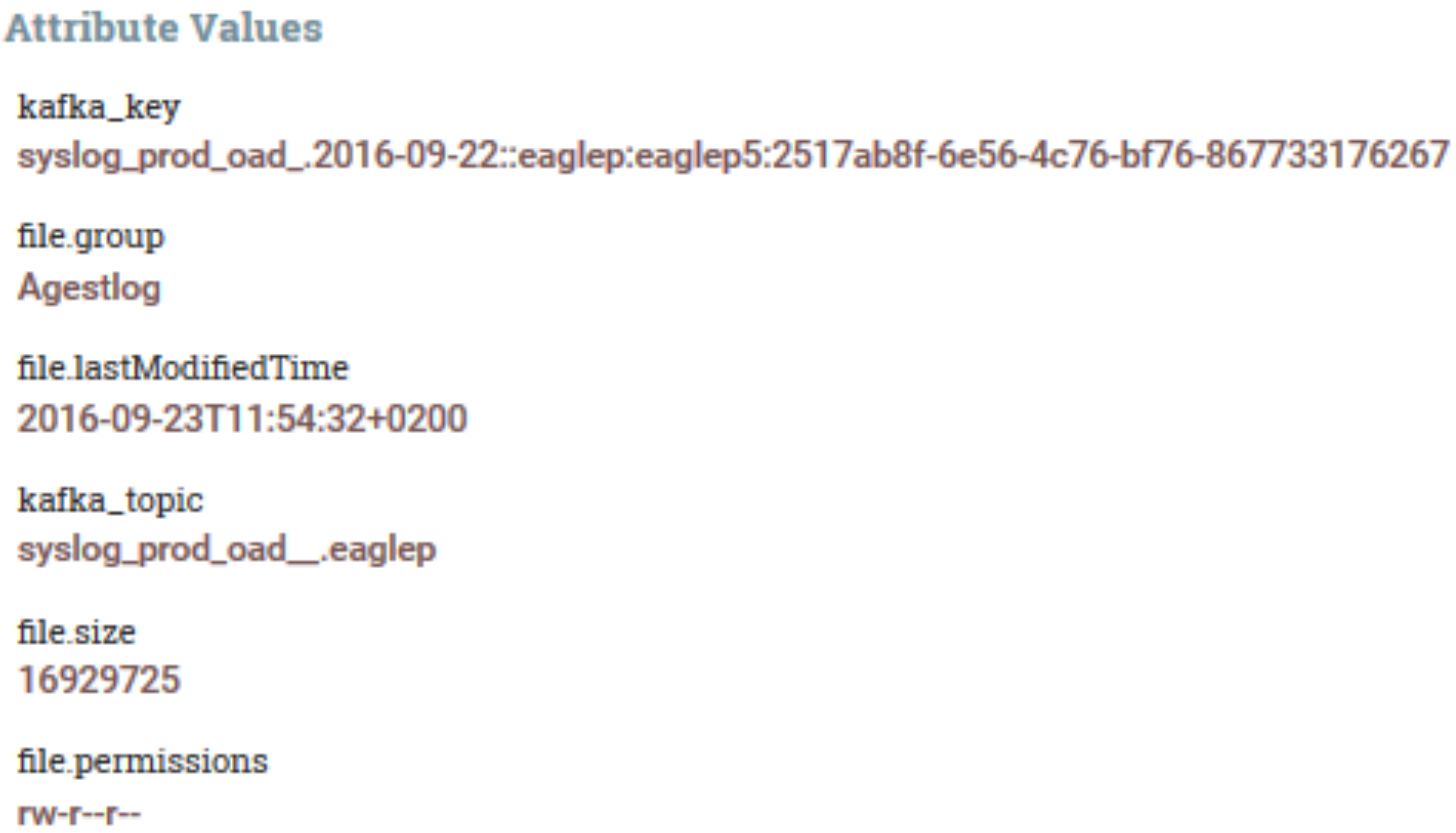
High level analytics

- Outliers detection
- SQL aggregations
- Timeseries non destructive sampling
- Query matching
- Network footprint clustering

Deployment


cluster Hadoop
Tableaux de bord, alertes et analyse
Kibana (+ grafana ?)
Elasticsearch indexe les évènements pour permettre leur recherche et leur agrégation en temps réel
Elasticsearch
es node
es node
es node
surveillance de /data/depot
Nifi Cluster
Nifi Master
Nifi Node
Nifi Node
Nifi Node
Indexation des évènements et alertes
(ré)indexation de données historiques
HDFS
data node
data node
data node
Calcul d'indicateurs pour la génération des alertes
Injection des logs bruts
Hadoop stocke durablement les évènements et permet le calcul d'indicateur sur de longues périodes
Analytique 360, alertes fil de l'eau et évènements rares
LogIsland
Worker Spark
Worker Spark
Worker Spark
Collecte des évènements
Kafka
broker
broker
broker
Kafka bufferise les événements pour permettre la production/consommation fenêtrée des messages en asynchrone et en parallèle
LogIsland est une plateforme de Complex Event Processing qui permet de lever des alertes dynamiquement

Message header



Attribute Values

kafka_key
syslog_prod_oad_.2016-09-22::eaglep:eaglep5:2517ab8f-6e56-4c76-bf76-867733176267

file.group
Agestlog

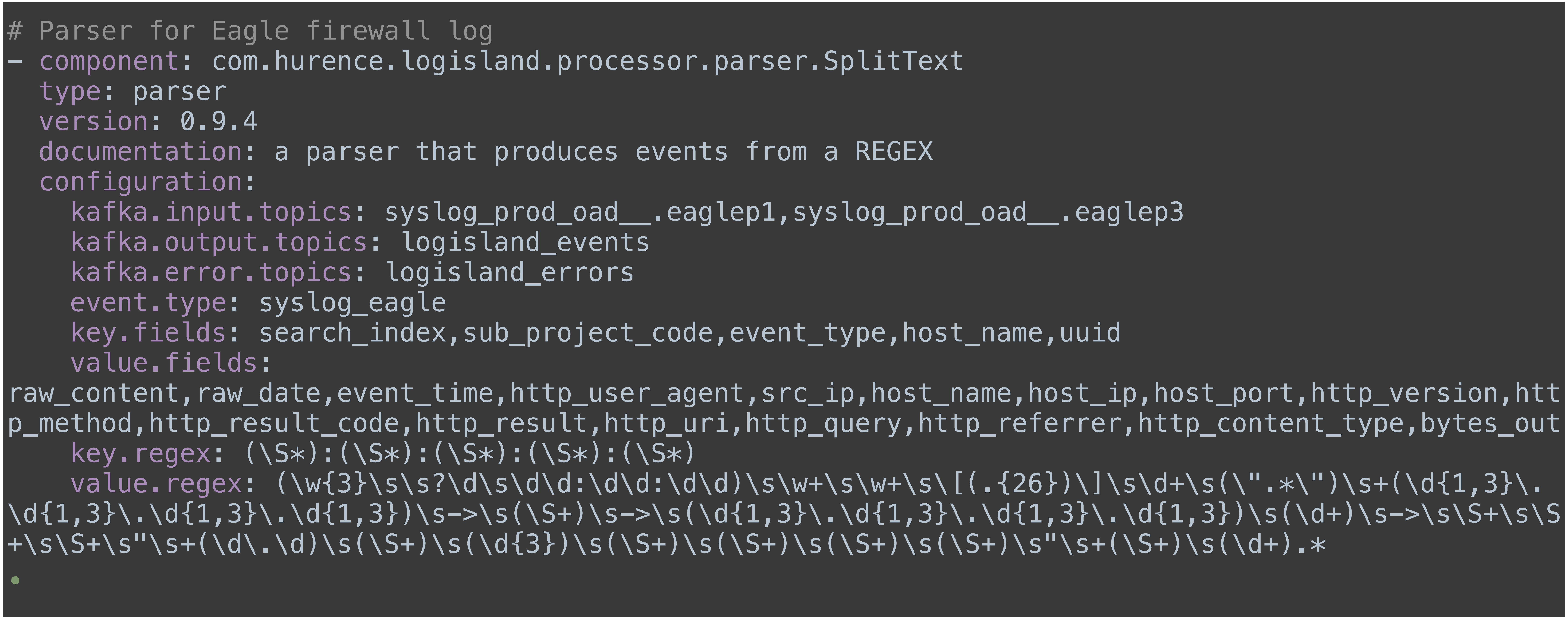
file.lastModifiedTime
2016-09-23T11:54:32+0200

kafka_topic
syslog_prod_oad__.eaglep

file.size
16929725

file.permissions
rw-r--r--

Configuration



```
# Parser for Eagle firewall log
- component: com.hurence.logisland.processor.parser.SplitText
  type: parser
  version: 0.9.4
  documentation: a parser that produces events from a REGEX
  configuration:
    kafka.input.topics: syslog_prod_oad__.eaglep1,syslog_prod_oad__.eaglep3
    kafka.output.topics: logisland_events
    kafka.error.topics: logisland_errors
    event.type: syslog_eagle
    key.fields: search_index,sub_project_code,event_type,host_name,uuid
    value.fields:
raw_content,raw_date,event_time,http_user_agent,src_ip,host_name,host_ip,host_port,http_version,htt
p_method,http_result_code,http_result,http_uri,http_query,http_referrer,http_content_type,bytes_out
    key.regex: (\S*):(\S*):(\S*):(\S*):(\S*)
    value.regex: (\w{3}\s\s?\d\s\d\d:\d\d:\d\d)\s\w+\s\w+\s\[(.{26})\]\s\d+\s(\".*\")\s+(\d{1,3}\.
\d{1,3}\.\d{1,3}\.\d{1,3})\s->\s(\S+)\s->\s(\d{1,3}\.\d{1,3}\.\d{1,3}\.\d{1,3})\s(\d+)\s->\s\S+\s\S
+\s\S+\s"\s+(\d\.\d)\s(\S+)\s(\d{3})\s(\S+)\s(\S+)\s(\S+)\s(\S+)\s"\s+(\S+)\s(\d+).*
```

Roadmap

- Auto-scaling for optimal cluster resource allocation
- Python processor with NLTK, Scrappy and Scikit libraries
- DNS tunneling detection
- Timeseries auto-regression forecasting
- Map/Matching with OpenStreetMap integration
- Enterprise frontend packaging
- Behavior clustering (host, user, …)

Why logisland ?

scalable

fast

high throughput

free

no code