

OpenSIPS and BigData

How to integrate OpenSIPS with ElasticSearch

Presenter

- Flavio E. Goncalves
 - Owner of the VOffice Group in Brazil
 - CTO of SipPulse Tecnologia Ltda.
 - OpenSIPS New Book and OpenSIPS New Bootcamp

SipPulse

- SipPulse is a Brazilian company dedicated to VoIP applications
 - SipPulse Anti Fraud System - TFPS
 - SipPulse Routing and Billing
 - Session Border Controllers
 - SIP-I/SIP-T Translators
 - Media Gateway Controllers
- More than 50 small to medium Telcos running SipPulse and OpenSIPS

The problems we were facing

- Logs from different sources
- Commands in the wrong console syndrome
- Time spent to get the information required to troubleshoot
- Logs stored only for a few days

Master Tool




Our industry generates a lot of data

- Billions of CDRs
- Terabytes of Traces
- Gigabytes of Logs
- Not easy to capture, transmit, store and search

BIG DATA

How can we make this data valuable?

1. Reduce the time to troubleshoot problems centralizing logs
 - 70% of the troubleshooting time is to collect data
 2. Enhance customer service quickly solving billing issues
 - Churn is a major problem in UCaaS and ITSPs
 3. Decreasing the calls to invalid or disconnected numbers
 - In some mailings more than 5% of the numbers are invalid
 4. Search numbers and IPs used for fraud and block real time
 - 20% of the numbers used in Toll Fraud are reused
 5. Discover patterns with analytics and better serve your customers
- 

ELK

- Elastic Search – Search and Analyze Data
 - Open Source Search Engine based on Apache Lucene
- Logstash – Process any data from any source
 - Open Source Log Contextualizer
- Kibana – Explore and Visualize
 - Open Source Analytics

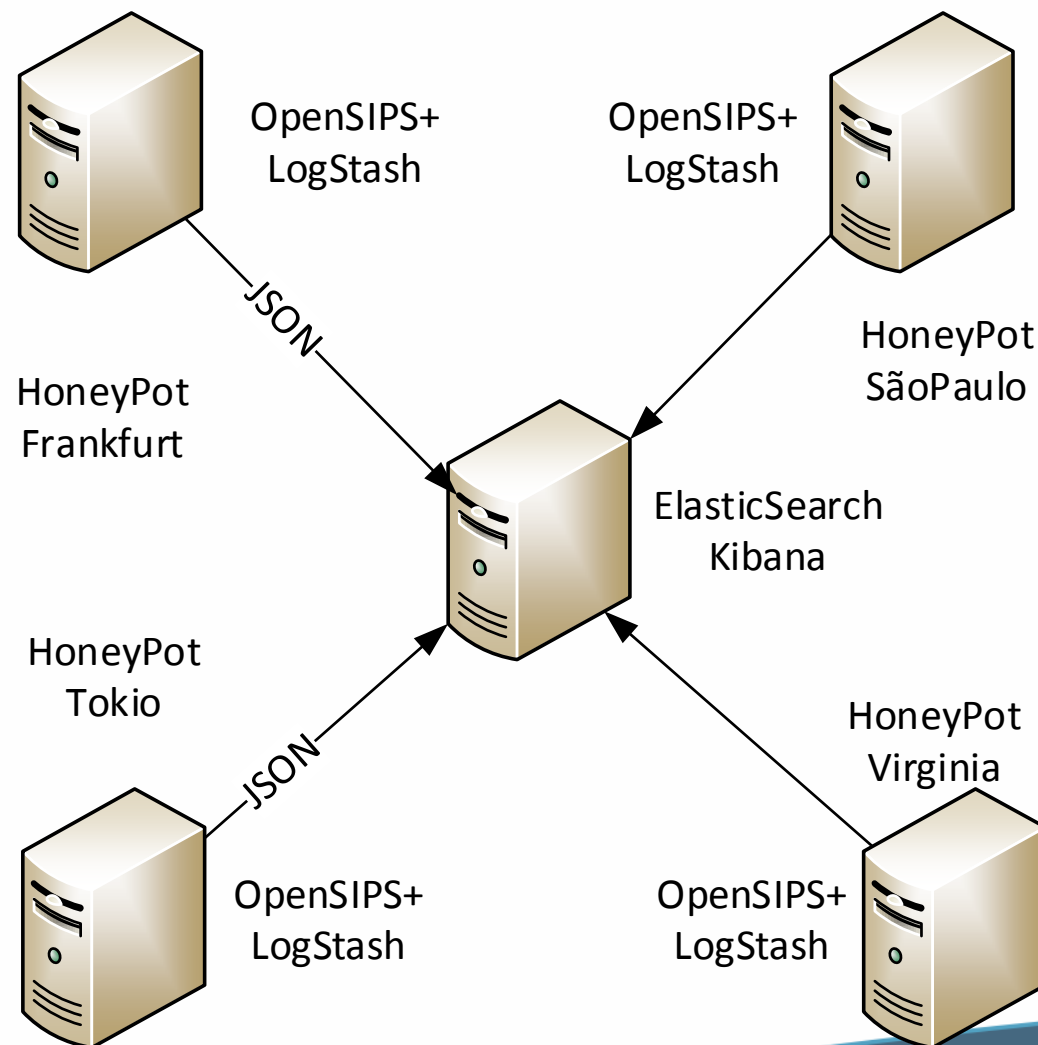


Case Study

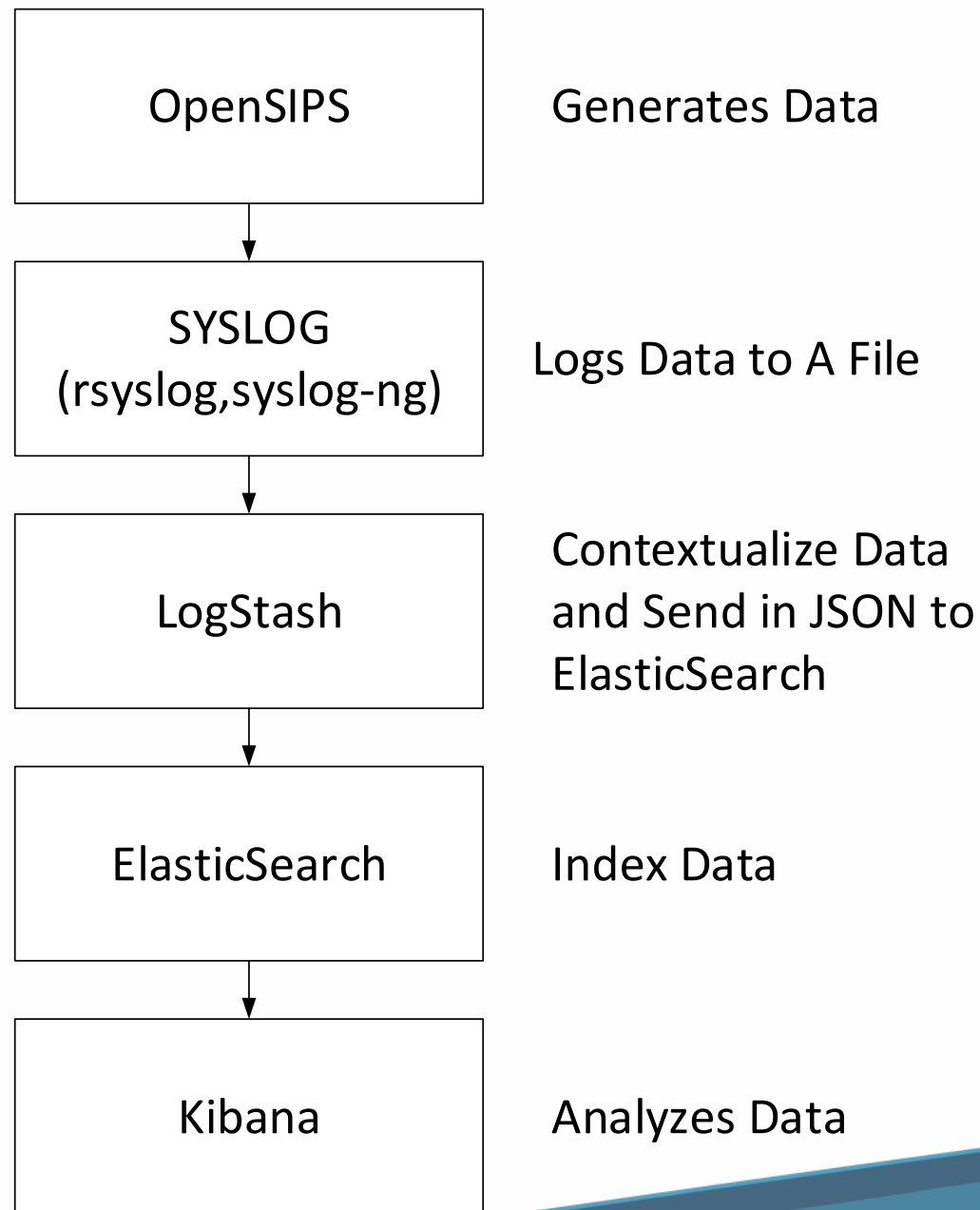
- Using ELK for Anti-Fraud Information
- Problems
 - Number formatting
 - 9011, 011, +, 901511.....
 - Quick access to online information
 - Search Numbers and IPs in real time
 - Provide easy access to information
 - Concerns regarding the delivery of information using MySQL over the Internet
 - Separate Databases for Online and Offline information

Recipe for a HoneyPot

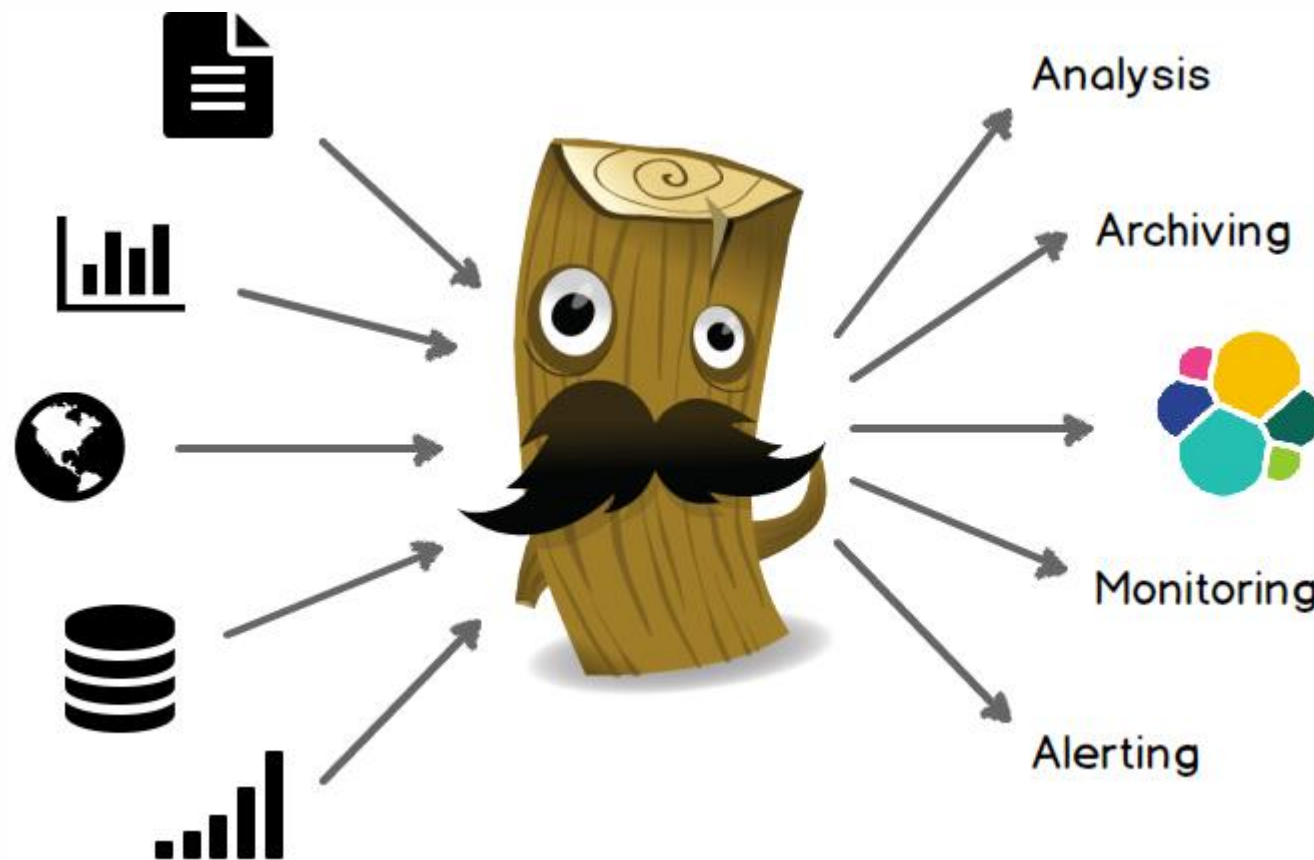
- OpenSIPS 2.1
- Apache
- Distributed DataCenter
- Frequent IP migrations



Data Flow Diagram



Logstash loves data!



<https://www.elastic.co/guide/en/logstash/current/introduction.html>

200 Available Plugins, No Plugins for OpenSIPS

GROK Is your friend!

Parse arbitrary text and structure it.



How GROK Works

<http://grokconstructor.appspot.com/do/match>

```
filter {
  grok {
    match => {
      "message" => "%{SYSLOGTIMESTAMP:syslog_timestamp}
%{DATA:syslog_program}(?:\[%{POSINT:syslog_pid}\])?:
%{IP:honey_pot_ip},%{IP:intruder_ip},%{WORD:sip_method},sip:%{WORD:ani}@%{HOSTNAME:ani_domain}
},sip:%{GREEDYDATA:dnis}@%{GREEDYDATA:dnis_domain},%{GREEDYDATA:user_agent},\[%{NUMBER:longitude},%{NUMBER:latitude}\]"
    }
  }
}
```



ElasticSearch

<http://w.x.y.z:5500>

```
{
  "name" : "Jonathan Richards",
  "cluster_name" : "elasticsearch",
  "version" : {
    "number" : "2.3.1",
    "build_hash" : "bd980929010aef404e7cb0843e61d0665269fc39",
    "build_timestamp" : "2016-04-04T12:25:05Z",
    "build_snapshot" : false,
    "lucene_version" : "5.5.0"
  },
  "tagline" : "You Know, for Search"
}
```

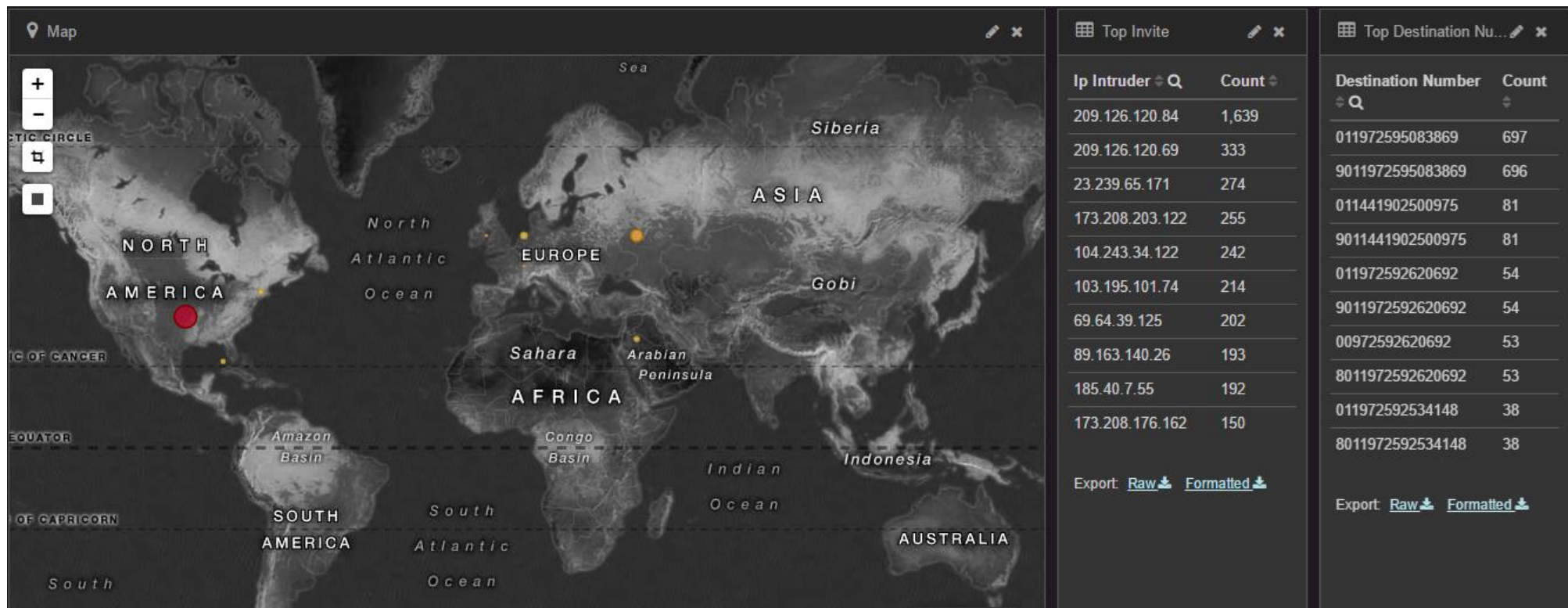
Searching

http://w.x.y.z:5500/_search?q=972598294121

http://w.x.y.z:5500/_search?q=friendly-scanner

http://w.x.y.z:5500/_search?q=173.208.203.122

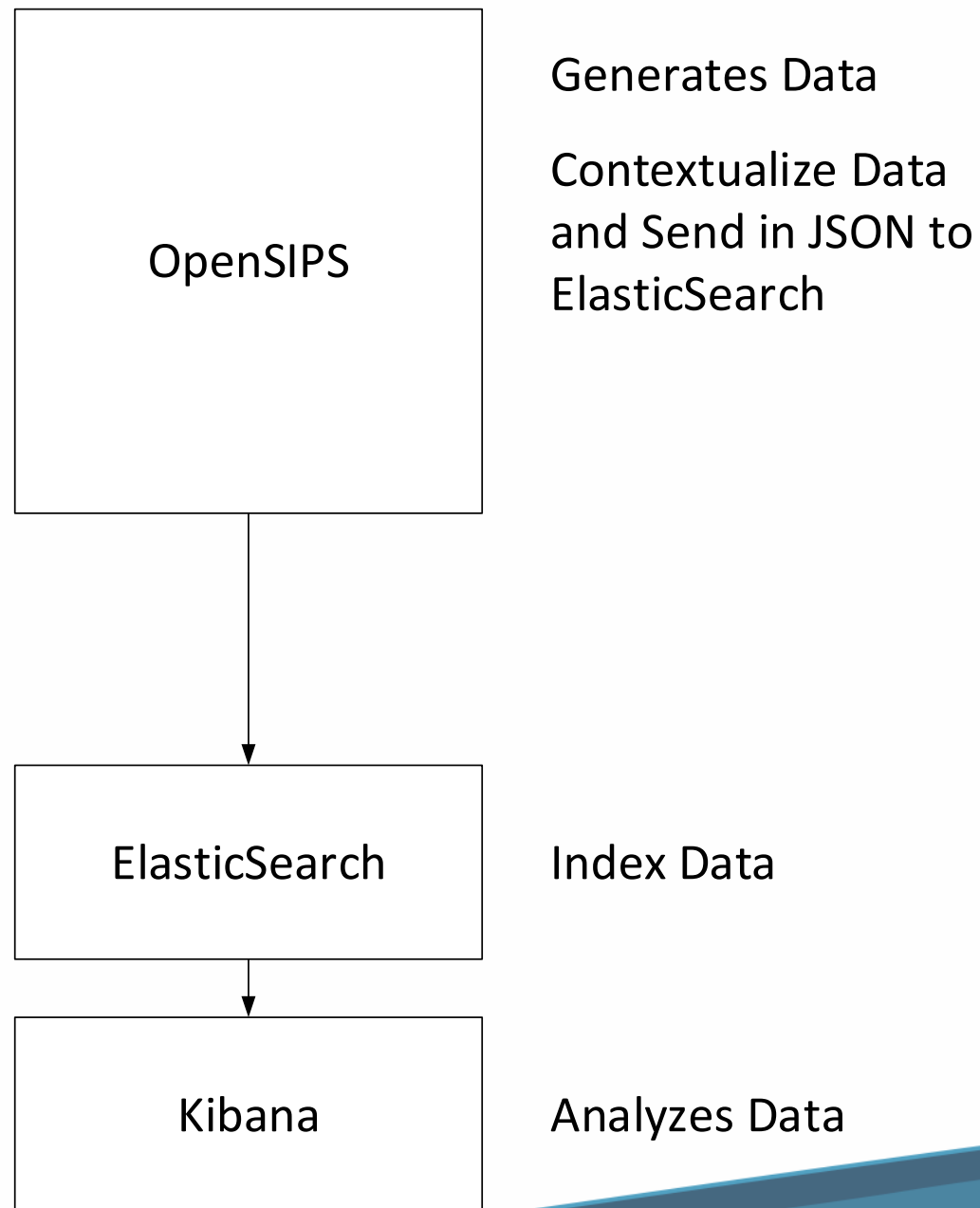
Kibana - Analytics



OpenSIPS Integration

- Logstash is based on Java and a bit slow
 - We can bypass Logstash sending data straight from OpenSIPS
 - We want also to consume data directly from Elastic Search
- 

New Data Flow



OpenSIPS Integration

```

if (is_method("INVITE")) {
    #####Create crud json
    $json(body) := "{}";
    $json(body/time) = $time(%F %T-0300);
    $json(body/sipRequest) = "INVITE";
    $json(body/ipIntruder) = $si;
    $json(body/destNum) = $rU;
    $json(body/userAgent) = $ua;
    $json(body/country)=$var(city);
    $json(body/location)=$var(latlon);
    $json(body/ipHost) = $Ri;

```

```

async(rest_post("http://user:password@w.x.y.z:9200/opensips/1", "$json(body)", "$var(ctype)",
"$var(ct)", "$var(rcode)"),resume)

```

Now OpenSIPS can go straight to the data!

```
if (rest_get("http://user:password@w.x.y.z:5500/_count?q=destNum:$rU&pretty",
"$var(body)", "$var(ctype)", "$var(rcode)")) {
    $json(body) := $var(body);
    if ($json(body/count) != 0) {
        xlog("Exists\n");
        exit;
    } else {
        xlog("Don't Exist\n");

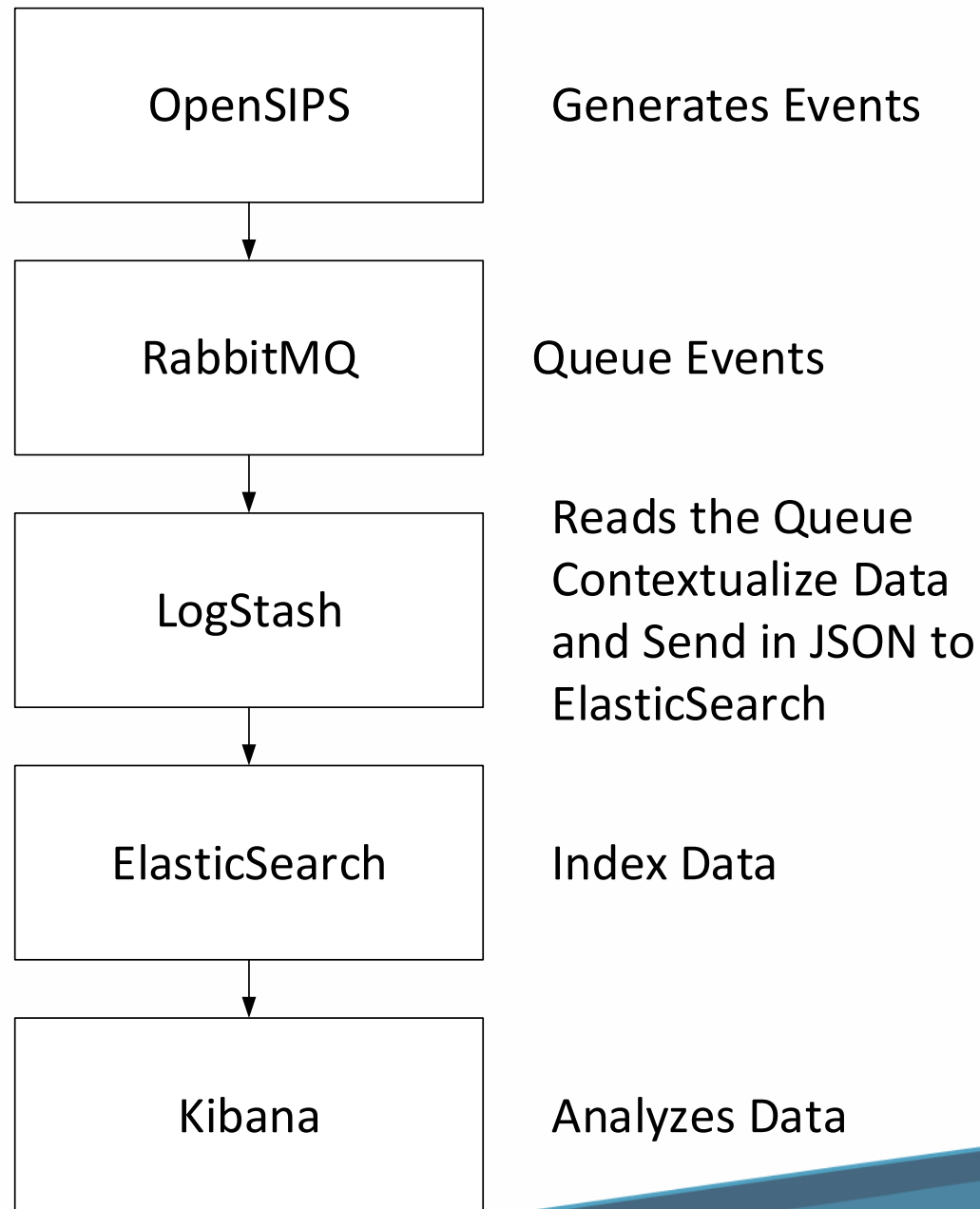
        # ...

    }
}
```

What else?

- Quick Billing Log (RabbitMQ Plugin)
 - INVITE REQUEST
 - INVITE REPLY
 - BYE REQUEST
 - BYE REPLY
 - DATA:
 - request_time,
 - reply_time,
 - caller_id,
 - callee_id,
 - call_id,
 - microseconds,
 - reply_code
- Purpose:
 - Resolve billing discrepancies without sending a ton of data over the Internet

Data Flow





Advantages of Elastic Search

- Free and Open Source
- Quick, easy and powerful search capabilities
- Unstructured and correlated data:
 - logs,
 - cdrs
 - and eventually traces (Homer can export)
- Control over the size of the data sent
- Less costly to store in AWS.
- Easy Analytics

OpenSIPS and ElasticSearch

- Integration via Syslog
- Integration via REST_CLIENT
- Async Calls have low effect on SIP server performance
- Several use cases:
 - Centralizing logs
 - Anti-Fraud
 - Do not call blacklists

Scalability

- Vertical Scalability
 - More powerful hardware is not always the solution
- Horizontal Scalability
 - Cluster Ready
- Data Center Services
 - AWS ElasticSearch
- HipChat
 - 1.2 Billion messages
 - 8 ElasticSearch Servers
 - 60 messages per second

Further Investigation

- SYSLOG-NG can be a good replacement for logstash
 - Developed in C is probably much faster than logstash
 - It is capable to send data straight to ElasticSearch

```
@module mod-java
@include "scl.conf"

destination d_elastic {
  elasticsearch(
    index("syslog-ng_${YEAR}.${MONTH}.${DAY}")
    type("test")
  );
};
```

Conclusion

- ElasticSearch seems to be a viable platform for big data and to handle Logs and CDRs.
- ElasticSearch can be integrated with OpenSIPS using the REST_CLIENT directly, RABBITMQ and SYSLOG in combination with Logstash.
- This is a preliminary research, so we are not aware yet of scalability problems of the model. Horizontal scalability helps, but the cost/benefit has to be measured compared to SQL and NoSQL approaches

QA

Contact Information

- E-mail: flavio@sippulse.com
- LinkedIn: <https://br.linkedin.com/in/flavioegoncalves>
- Twitter: #flagonc