

**Analyzing large flow data sets using
modern open-source data search and
visualization tools**

FloCon 2014

Max Putas

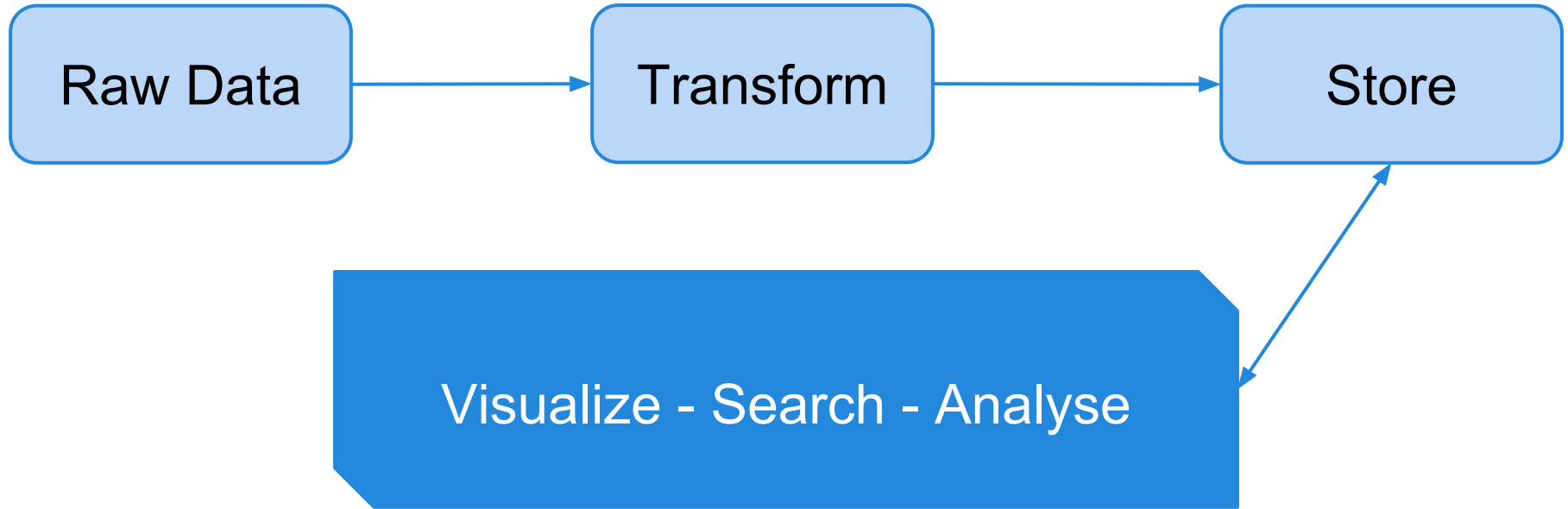
About me

- Operations Engineer - “DevOps”
- BS, MS, and CAS in Telecommunications
- Work/research interests
 - System automation
 - Efficiency improvement
 - System and network monitoring
 - Traffic/service analysis
 - Open-Source software

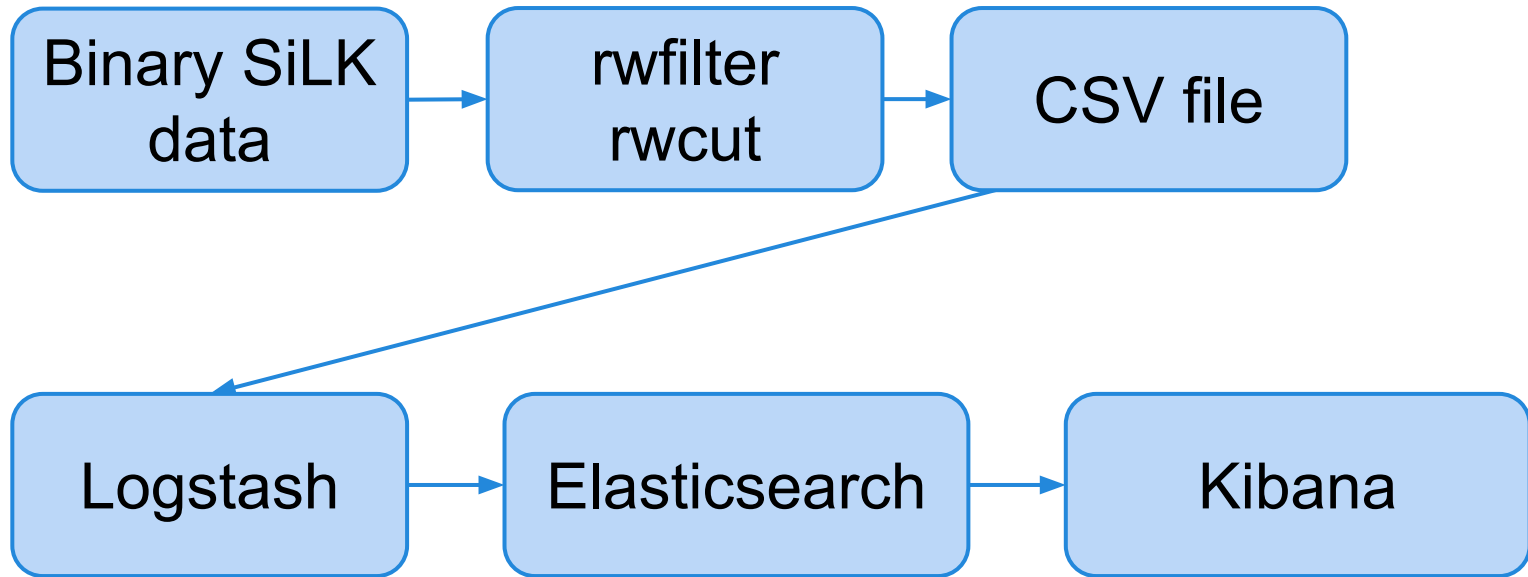
Common tools for analysis

- **Scripts: Bash, Perl, Python**
 - Learning curve, time-intensive
 - GnuPlot for graphing/visualization
- **Application-specific tools**
 - SiLK, Apache Chainsaw, Wireshark
- **Splunk - EXPENSIVE**
- **Excel**
 - :-)

General model



Components



Logstash : About

- Can act as an agent, server, or both
- Single jar file – only depends on Java
- Very young project
 - Started in late 2010
 - First official book released last year (2013)

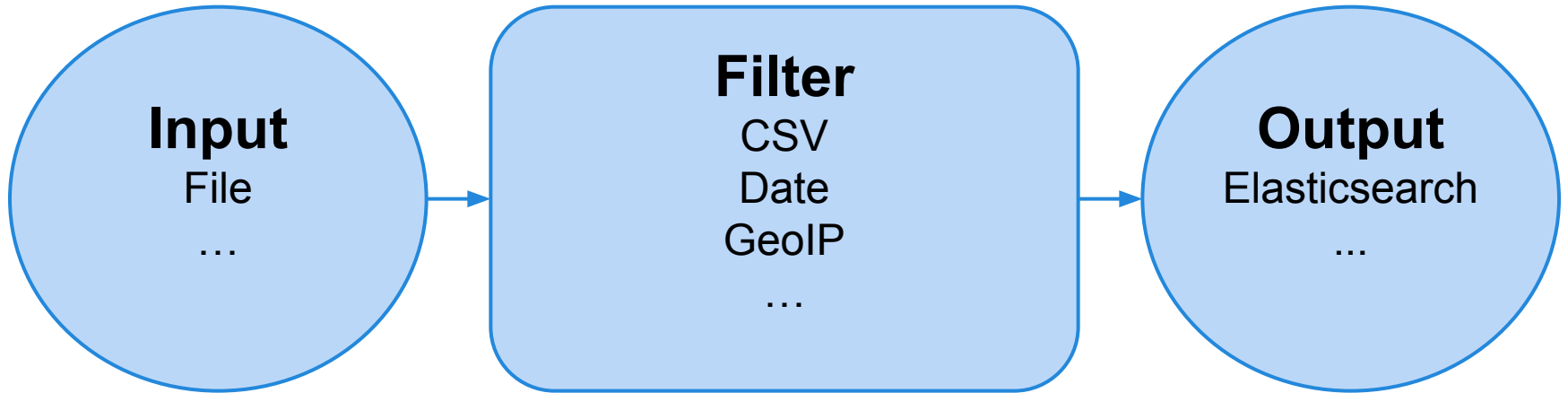


Logstash : Plugins

36

40

46



Logstash : Configuration

```
input {
  file {
    path => "/tmp/silk-data.csv"
    start_position => "beginning"
    type => "silkcsv"
  }
}
filter {
  ...
  date {
    type => "silkcsv"
    match => [ "sTime", "yyyy/MM/dd'T'HH:mm:ss.SSS" ]
    add_tag => [ "dated" ]
  }
  ...
}
output {
  elasticsearch { host => "localhost" }
}
```

Elasticsearch : About

- Built on Apache Lucene (indexing/search library)
- Java
- RESTful API
- Distributed, scalable architecture.
 - Nodes can find each other through discovery
- JSON-based
- "Big data" focus

elasticsearch.

Elasticsearch : Data storage

- **Index** - document “database”
 - Document types → fields → type mappings
- **Shards** - pieces of the index
 - More shards, better indexing performance across the cluster
- **Replicas** - how many copies of each shard
 - More replicas, better search performance and redundancy



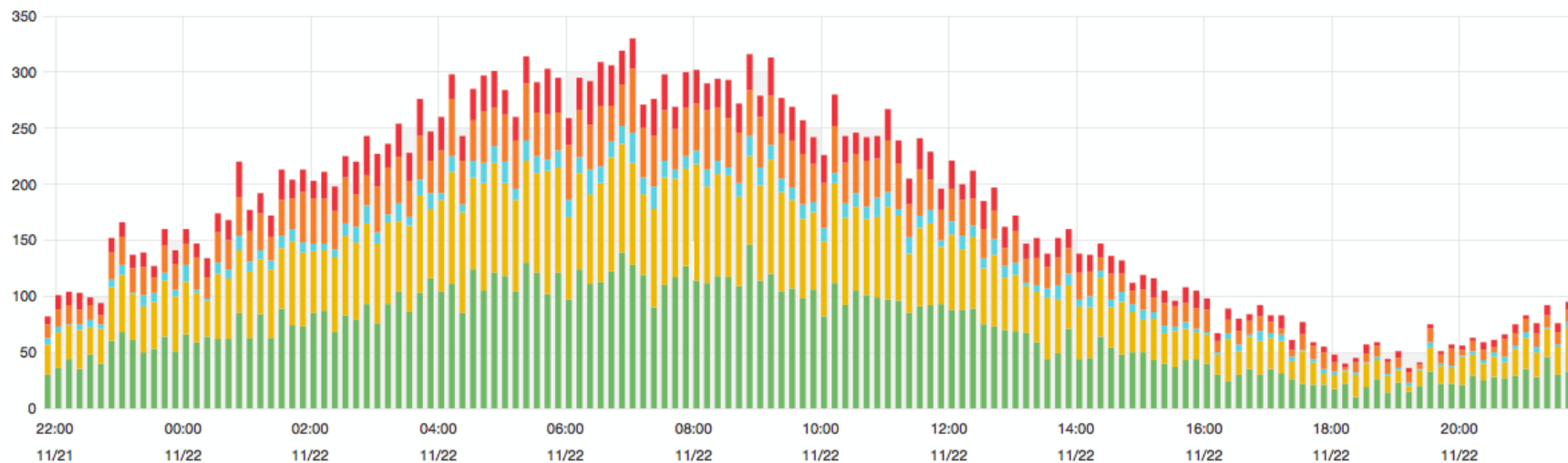
Elasticsearch : Performance

- Lab setup: 6-core CPU : 16GB RAM : SATA HD
 - Indexing performance: 4000/s
- Double the number of shards and machines
 - ~2x index performance increase
- Double the number of replicas
 - ~2x search performance increase
- Can take full advantage of SSDs

Elasticsearch : Type Mapping

```
...
"dIP" : {
  "type" : "ip"
},
"dPort" : {
  "type" : "integer"
},
...
"duration" : {
  "type" : "float"
},
...
"eTime" : {
  "type" : "date",
  "format" : "yyyy/MM/dd'T'HH:mm:ss.SSS"
},
...
```

Kibana



Kibana : Features

- **Pure Javascript**: connects directly to Elasticsearch
 - A reverse proxy will be necessary to limit access
- **Graphing/visualization**: histograms, scatter plots, pie charts, ranked lists, maps, and line graphs
- **Statistics**: trends, min, mean, and max
- **Real-time search**: Simultaneous queries, sortable results, filters, field drill-down, and derived (faceted) queries

Development

- The developers of Kibana and Logstash were recently hired by Elasticsearch

elasticsearch.



Kibana

- There is a possibility of even tighter integration in the future

More possibilities

- Logs
 - Web, database, e-mail, and DNS servers
 - Firewalls, IDS/IPS, switches, and routers
 - Syslog and Windows events
- Monitoring alerts: SNMP
- Performance metrics
- Others?
 - If it's textual and log-like it'll probably work
 - Custom plugins are possible
- Gather related data to correlate events in Kibana or through the Elasticsearch API

Parsing

- Logstash provides built-in parsing (“grok”) rules:

```
HTTPDATE %{MONTHDAY}/%{MONTH}/%{YEAR}:%{TIME} %{INT}
```

- Common Apache log format:

```
127.0.0.1 - frank [10/Oct/2000:13:55:36 -0700] "GET  
/apache_pb.gif HTTP/1.0" 200 2326
```

- Complete rule:

```
APACHELOG %{IPORHOST:clientip} %{USER:ident} %{USER:auth}  
\[%{HTTPDATE:timestamp}\] "(?:%{WORD:verb} %{NOTSPACE:  
request} (?:HTTP/%{NUMBER:httpversion})?|%{DATA:  
rawrequest})" %{NUMBER:response} (?:%{NUMBER:bytes}|-)
```

DEMO

References and resources

- <http://logstash.net>
- <http://www.elasticsearch.org>
- <http://www.elasticsearch.org/overview/kibana/>
- Try Kibana yourself:
 - <http://demo.kibana.org>
- Debug grok parsing rules:
 - <http://grokdebug.herokuapp.com>
- SiLK Kibana 3 demo video:
 - <https://vimeo.com/71393353>
- Contact: max.putas@gmail.com

