Scalable logging and log-file analysis

High-Performance Computing System Administration

Linus Weber

Institute of Computer Science Göttingen

February 2023

Log-file analysis

- Similar to monitoring (TIG stack), but instead of collecting metrics (time-series data) it is about collecting and analyzing log-files (or log-entries).
- Store collected files on centralized server.
- Variety of possible input sources.
- Agents - Collector - Storage - Visualization & Analysis

Scalable logging

Here: scale the level of detail, not the capacity of the infrastructure.

- File selection
- Log level (debug, info, warning, error)
- Filters
- Aggregation
- Metrics polling interval
- Retention period

Why do scalable logging?

- Different environments (dev, test, staging, production)
- New instance
- System update
- Traffic peaks
- Unexpected failure

Use-cases require custom-tailored scaling solutions.

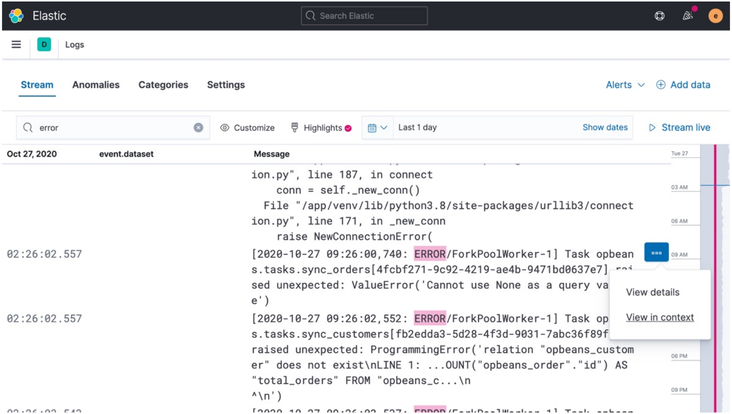
Types of services

1. Provider-managed
2. Software-as-a-Service: Google Cloud logging, Sumo Logic, Loggly
3. Self-hosted: Elastic Stack, Icinga, Nagios Log Server, Graylog, Splunk (available as SaaS)
4. Custom solution using applications developed in-house and components like Redis, Kafka, Elasticsearch, InfluxDB, SaaS, etc.

Elastic Stack

- Agent - Beats
- Collector - Logstash
- Storage - Elasticsearch
- Visualization & Analysis - Kibana

Kibana Logs



```
Oct 27, 2020    event.dataset    Message
                                 ion.py", line 187, in connect
                                     conn = self._new_conn()
                                   File "/app/venv/lib/python3.8/site-packages/urllib3/connect
                                 ion.py", line 171, in _new_conn
                                     raise NewConnectionError(
02:26:02.557                     [2020-10-27 09:26:00,740: ERROR/ForkPoolWorker-1] Task opbean
                                 s.tasks.sync_orders[4fcbf271-9c92-4219-ae4b-9471bd0637e7] rai
                                 sed unexpected: ValueError('Cannot use None as a query va
                                 e')
02:26:02.557                     [2020-10-27 09:26:02,552: ERROR/ForkPoolWorker-1] Task op
                                 s.tasks.sync_customers[fb2edda3-5d28-4f3d-9031-7abc36f89f
                                 raised unexpected: ProgrammingError('relation "opbeans_custom
                                 er" does not exist\nLINE 1: ...OUNT("opbeans_order"."id") AS
                                 "total_orders" FROM "opbeans_c...\n
                                 ^\n')
```

Figure: View logs in Kibana. Source: https://www.elastic.co/de/kibana/

Kibana Charts

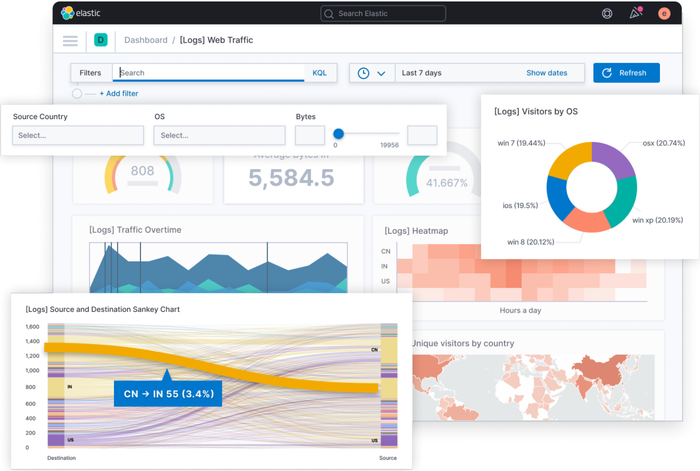


Figure: Various chart types in Kibana. Source: https://www.elastic.co/de/kibana/

Challenges

- Requirements engineering
- Make it scale
- Integrate with existing systems