EUNIS 2019: Log analytics using ELK stack on Cloud platform

1 Jordi Cuní SIGMA AIE (SPAIN)
2 Estefania Muñoz SIGMA AIE (SPAIN)

Keywords
Lucene, Elastic search, logstash, kibana, audit,monitoring

1. ABSTRACT

Elasticsearch is an open-source, RESTful, distributed search and analytics engine built on Apache Lucene. Elasticsearch is typically used for loging analytics, full-text searching, security intelligence, business analytics and operational intelligence use cases. Logstash is a tool to collect, index, forward events and log messages. Kibana is an open source data visualization plugin for Elasticsearch. It provides visualization capabilities on top of the content indexed on an Elasticsearch cluster.

In SIGMA AIE [2] we use this architecture for writing useful information about how the users are using our application and also build dashboards in ELK with online information. We are capable to know how many users are logged in, how long they are connected, what are they request to the application, and if in the GDPR environment they request personal data. All this audits aspects have been solved throught the analisys of the applications logs. Spending little time and effort we have turned our system with an audit layer very necessary for nowadays.

2. THE PROJECT

SIGMA Gestion Universitaria [3] is a nonprofit consortium established in 1996 by a group of 8 top level Spanish Public Universities to provide technological solutions to their needs for managing academics, learning, research and organization processes. SIGMA represents 20% of the students in the Spanish university system. The consortium's objective has evolved towards the continuous technological modernization of university management through the development of IT solutions aimed at automating the administrative processes and, as a result, guaranteeing their effectiveness.

SIGMA has been working to collect data from the user to know how they interact with the application, what are they doing, how many time spend in the application, not only for the data analysis that can be done, but also to cover legal issues like the GDPR, that tell us to registry any activity related with the personal data.

The logs of the application usually are written with useful information for the developers in order to follow the code when is running to solve issues.We have use this well proven functionality to add user information about the application and what is done with it, not in a technical point of view but with usage point of view.

So the project was focused in define a log trace and all the developer teams add this new traces in their modules. This was a very simple action in time and effort. Each of the teams are including in their modules the new log traces and when they publish the new releases the new traces are collected by the ELK stack. For our organization this solution has been very transparent for the development cycle.

The following points are aspects of the GDPR and how ELK can allow to cover them [4]

- **Data Flow Mapping**: Mapping data flows is the first step in GDPR preparation, and if an organization is unable to identify relevant data flows, the GDPR initiative may be incomplete. Depending on where an organization is storing Personal Data, it may make sense to index information about the data flow into Elasticsearch, where its powerful and fast full-text search capabilities will enable quick identification of tables, queries, reports, or applications that rely on Personal Data.
- **Personal Data Retention Planning**: GDPR specifies limited retention, and GDPR-affected organizations are required to delete Personal Data when it is no longer needed (or when the Data Subject withdraws consent). The retention of Personal Data stored in Elasticsearch can be easily managed through index management. Elasticsearch supports time-based indices — that can be deleted after the retention period has expired
- **Access Controls**: To prevent unauthorized access to Personal Data stored in an Elasticsearch cluster, there must be a way to authenticate users. Elastic security features are able to integrate with those systems to perform user authentication. Elastic security features also include IP-based filtering.
- **Disaster Recovery**: Elasticsearch has been designed to be a distributed data store and search engine.Elastic scales horitontaly, this means that its own distributed architecture makes easy to recover from a disaster. And also elastic security features help to preserve the integrity of data with message authentication and SSL/TLS encryption.

Elastic Stack can be used as a centralized logging system to implement the GDPR principles of protection by design, cryptography and pseudonymization, plus the technical and organizational measures for protection of Personal Data, including access controls, logging and auditing, as well as monitoring and detection that can help to GDPR compliance.

3. FUTURE ACTIONS

We will continue adding more information in the logs and in the future we will build dashboard for the different modules available for our universities in Cloud mode.

4. REFERENCES

[1] ELK platform: https://www.elastic.co/

[2] SIGMA AIE http://soporte.gestionuniversitariasigma.com/index.php/es/

[3] SiS Student information System SIGMA ACADEMIC

[4] ELK White paper https://www.elastic.co/pdf/white-paper-of-gdpr-compliance-with-elastic-and-the-elastic-stack.pdf

5. AUTHORS' BIOGRAPHIES

Jordi Cuní

Chief Information Officer

Jordi was born in 1976 in Barcelona, Spain. He holds a Computer Science degree at Universitat Oberta de Catalunya (2006 – 2012) and a Computer and Software Engineer at Universitat Autònoma de Barcelona (1997 – 2000). He works at SIGMA since 2000, being the current Manager of the Architecture and Software quality assurance Areas. He leads a team of 7 developers in those areas. His role mainly focuses on the maintenance and development of the (own) Sigma framework to increase the productivity, define the methodology among the different areas and establish the software development tools for the rest of the company. Last but not least, his area also takes part on the technical and performance customer support helping the clients with the migration projects of their back-end resources.

Previously, he had been project manager developing SIGMA's educational planning area for 5 years. His main efforts focused on the development of software solutions, resource planning, stock management of static and mobile resources and offline and on-line surveys.