27 September 2021

Alforat.al-sakkaf@cern.ch

CERN summer student programmer 2021.

'Centralized Elasticseach' team at CERN.

Author : Alforat Alsakkaf

Supervisors : Sokratis Papadopoulos

CERN – European Organization for Nuclear Research,
Geneva-Switzerland.

Contents

1-Introduction

After completing the summer internship I made this report as part of my summer project during CERN Summer Program 2021, being from a non-member country.

I was part of the elasticsearch team where we configure the machines, run the service and help our users with any problem they may have.

My project is focused on using a side-product of Elasticsearch, Kibana, which is also developed by Elastic. Kibana is a user interface that lets you visualize your Elasticsearch data. The "equivalent" OpenSearch project, is named "OpenSearch Dashboards". Kibana comes with different types of visualizations (bar chart, line chart, etc.) and allows the creation of other "custom" visualizations as external plugins that can be used within Kibana. The team at CERN has developed the “Relational filter”.

As the team transitioned from Elasticsearch to OpenSearch and consequently from Kibana to OpenSearch Dashboards, I worked on linking the relational filter plugin that was developed for Kibana to the new OpenSearch Dashboards product. After finishing the project, the plugin can be contributed to the community of the OpenSearch project.

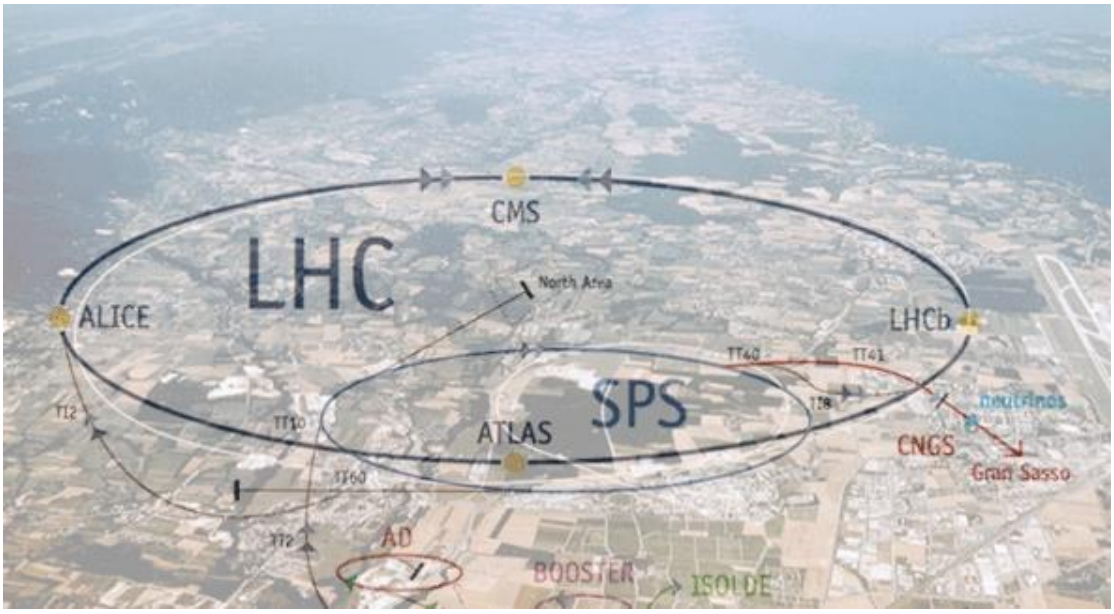
2-Background

This section covers an introduction to the CERN organization, as well as training for students, graduates and other aspirations.

2.1- About CERN

CERN was built at the end of the Second World War by a handful of visionary scientists that envisioned a European atomic physics laboratory. Raoul Dautry, Pierre Auger and Lew Kowarski in France, Edoardo Amaldi in Italy and Niels Bohr in Denmark were among these pioneers. French physicist Louis de Broglie put forward the first official proposal for the creation of a European laboratory at the European Cultural Conference, which opened in Lausanne on 9 December 1949. However only two years after, at an intergovernmental meeting of UNESCO in Paris in December 1951, the first resolution concerning the establishment of CERN was adopted. Two months later, 11 countries signed an agreement establishing the provisional council – CERN was born.

CERN is located exactly on the border between Switzerland and France. While the majority of offices and tertiary buildings are in Switzerland, the biggest part of the SPS and LHC accelerator is built in the French underground.



CERN employs around 2500 people (including the DG) from the 22 Member States but in total almost 11000 scientists of more than 100 different nationalities use CERN's infrastructure.

The Large Hadron Collider (LHC) took about a decade to build, at a total cost of about 4,322 million Swiss francs. There are many different experiments working at the LHC, including the CMS and ATLAS Detectors that discovered the Higgs boson. CERN contributes about 20% of the cost of those trials, bringing the total cost to about 5,300 M.

The LHC's total operating budget is about 1 billion Swiss francs annually, with funding from member states. It goes for operation, maintenance, technical downtime, repairs and consolidation work in people and materials.

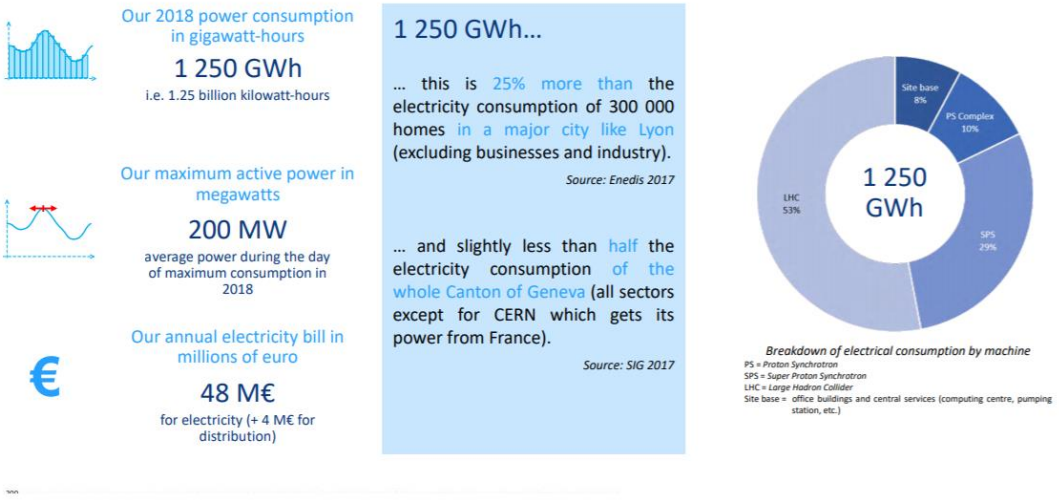
You might think, wow that's a lot. But at the end of the day, there is one or less one cappuccino for every European. Do you think it is worth it? I totally think so.

CERN's energy consumption The total energy consumption in CERN is 1.25 TWh per year while the total electrical energy production in the world is about 20,000 TWh, in the European

Union it is 3400 TWh, in France it is about 500 TWh, and in the Geneva Canton 3 TWh.

TWh of electricity per year is enough to power 300,000 homes for a year in Switzerland, or about a third of the energy used by neighboring Geneva in Switzerland. 200 MW instead represents the average power during the day of maximum consumption in 2018.

The annual electricity bill is around €48 million for electricity (+ €4 million for distribution). CERN receives its electricity via two high-voltage (400 kV) power lines connected to the French national grid at the Bois-Tollot substation near the Prévessin site. An emergency backup supply source (130 kV) provided by the Swiss network can be used in the event of a 400 kV supply failure, for example to operate and reconnect CERN public services and security systems (buildings, offices, safety systems, etc.).



Amount of data generated by CERN: the LHC experiments represent about 150 million sensors that deliver 30 million data entries per second.

This results in more than 50 petabytes each year, saving a total of 500 petabytes of data stored on disk and another 400 petabytes on tape. This is a very difficult task and will become even more difficult during the coming years when the transmission rate is expected to double due to the increased beam brightness pursued by the HL-LHC project. All of this data is stored in the CERN data center where approximately 260,000 processor cores and 15,000 servers operate 24/7. Data is also shared around the world through the Worldwide LHC Computing Grid, a massive network backed by 170 computing centers and 900,000 computing centers, which runs about two million missions per day and allows more than 12,000 physicists to access LHC data [1].

3-Time plan

Tasks	First week	Second week	Third week	Fourth week	Fifth week	sixth week	Seven week	Eighth week
Knowing the nature of the work								
Connecting to the development machine								
Setting up the development environment for Kibana								
Plugin development								
Setting up a new development node								

Note: Project stopped at the sixth week for special reasons and resumed after the eighth week.

4-Concepts

4.1- Elasticsearch is a distributed, free and open search and analytics engine for all types of data, including textual, numerical, geospatial, structured, and unstructured. Elasticsearch is built on Apache Lucene and was first released in 2010 by Elastic N.V. (now known as Elastic). Known for its simple REST APIs, distributed nature, speed, and scalability, Elasticsearch is the central component of the Elastic Stack, a set of free and open tools for data ingestion, enrichment, storage, analysis and visualization. Commonly referred to as the ELK Stack (after Elasticsearch, Logstash, and Kibana), the Elastic Stack also includes a rich collection of lightweight shipping agents known as Beats for sending data to Elasticsearch [2].

4.2- Open Distro for Elasticsearch aka OpenSearch project is well-suited to the following use cases [3]:

- Log analytics
- Real-time application monitoring
- Clickstream analytics
- Search backend

4.3- Kibana is an free and open frontend application that sits on top of the Elastic Stack, providing search and data visualization capabilities for data indexed in Elasticsearch. Commonly known as the charting tool for the Elastic Stack (previously referred to as the ELK Stack after Elasticsearch, Logstash, and Kibana), Kibana also acts as the user interface for monitoring, managing and securing an Elastic Stack cluster — as well as the centralized hub for built-in solutions developed within the Elastic Stack. Developed in 2013 from within the Elasticsearch community, Kibana has grown to become the window into the Elastic Stack itself, offering a portal for users and companies [4].

5-Tools

When I worked on my project, it is true, I did not finish it, but I accomplished most of it, completed many things, and used different tools, most of which were new to me. The most prominent tools used were to work remotely on a CERN OpenShift OKD, using lxplus service and SSH, tunneling. They played a great role and took a lot of time in order to be able to. Also the use of git and Gitlab.

Most notable was JavaScript, which played a major role. As for the collaborative and chatting tools that helped us communicate with my colleagues, it also enabled us to attend lessons remotely, the most prominent of which was the Zoom program, CodiMD for editing notes and configuration files and Mattermost for communicating with team members for coordinating activities.

6-Explanation

The work scenario was to connect to the development machine through *eskibanadev02* using Kerberos, then creating the development environment for Kibana with attention to the release.

The development plugin that we want to migrate is located at: https://github.com/cernops/kibana-relational-filter

A new development node is needed to create a new VM in OpenStack.

login as root (example: eskibanadev02)

connect to CERN (aiadm - preferred, requires 2FA, or lxplus)

ssh root@eskibanadev02.

After running the latest reboot kernel we configure the node

locmap --enable ssh

locmap --enable kerberos

locmap --enable afs

locmap --configure ssh

locmap --configure kerberos

locmap -- configure afs

Followed by adding the user and adding local working directories.

7-Resources and references

[1] https://meroli.web.cern.ch/blog_cern_questions.html

[2] https://www.elastic.co/what-is/elasticsearch

[3] https://opendistro.github.io/for-elasticsearch-docs

[4] https://www.elastic.co/what-is/kibana