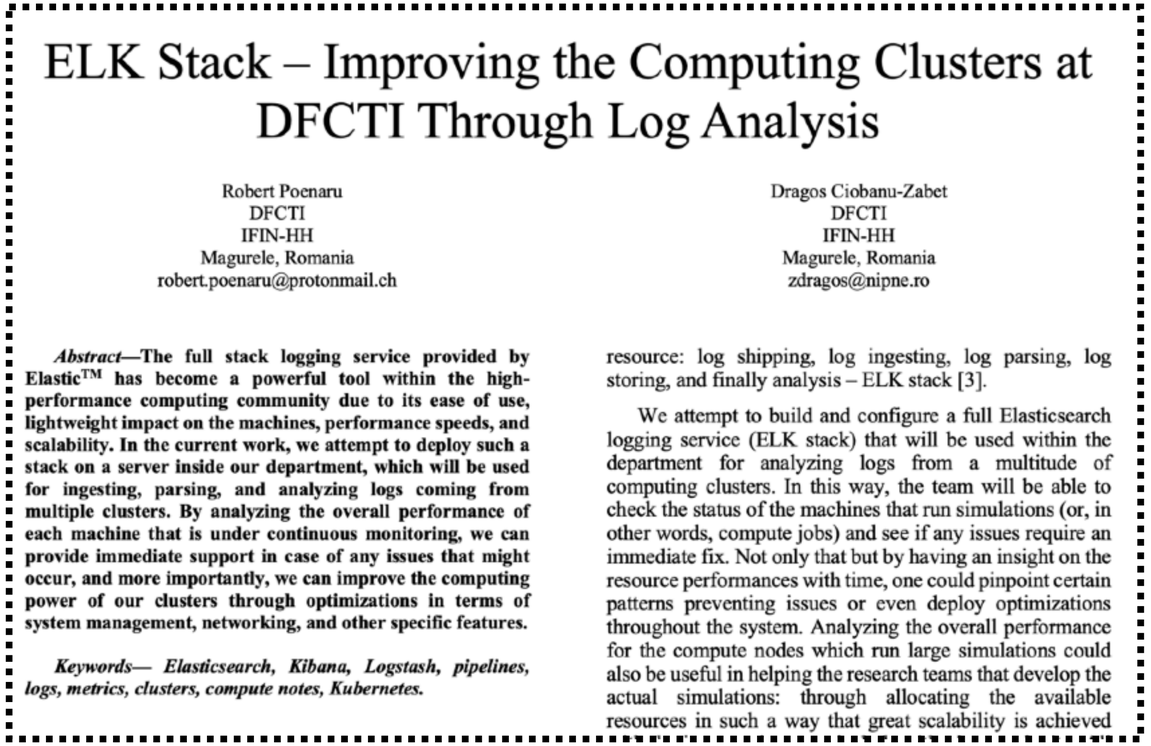
ELK Stack

Improving the Computing Clusters at DFCTI Through Log Analysis

Robert Poenaru | #roedunet #December #2020

Outline

- Aim & Motivation
- Elasticsearch + Logstash + Kibana
- Log filtering
- Parsing logs from multiple sources
- Improvements & Future plans



ELK Stack – Improving the Computing Clusters at DFCTI Through Log Analysis

Robert Poenaru
DFCTI
IFIN-HH
Magurele, Romania
robert.poenaru@protonmail.ch

Dragos Ciobanu-Zabet
DFCTI
IFIN-HH
Magurele, Romania
zdragos@nipne.ro

***Abstract*—The full stack logging service provided by Elastic™ has become a powerful tool within the high-performance computing community due to its ease of use, lightweight impact on the machines, performance speeds, and scalability. In the current work, we attempt to deploy such a stack on a server inside our department, which will be used for ingesting, parsing, and analyzing logs coming from multiple clusters. By analyzing the overall performance of each machine that is under continuous monitoring, we can provide immediate support in case of any issues that might occur, and more importantly, we can improve the computing power of our clusters through optimizations in terms of system management, networking, and other specific features.**

Keywords*— *Elasticsearch, Kibana, Logstash, pipelines, logs, metrics, clusters, compute notes, Kubernetes.

resource: log shipping, log ingesting, log parsing, log storing, and finally analysis – ELK stack [3].

We attempt to build and configure a full Elasticsearch logging service (ELK stack) that will be used within the department for analyzing logs from a multitude of computing clusters. In this way, the team will be able to check the status of the machines that run simulations (or, in other words, compute jobs) and see if any issues require an immediate fix. Not only that but by having an insight on the resource performances with time, one could pinpoint certain patterns preventing issues or even deploy optimizations throughout the system. Analyzing the overall performance for the compute nodes which run large simulations could also be useful in helping the research teams that develop the actual simulations: through allocating the available resources in such a way that great scalability is achieved