Title of the Project

Create and implement a Data model to standardize Kubernetes logs.

Possible Mentors

Primary Mentor : Peter Portante <pportant@redhat.com>, Tushar Katarki <tkatarki@redhat.com>
Co-Mentor : Miguel Perez Colino <mperezco@redhat.com>

Synopsis / Project Summary

This project aims to build and implement a Data model for logs in a large kubernetes cluster to process, correlate and query to make the troubleshooting easier and can reduce the time in finding root causes.

The objective of this proposal is to twofold.
1. Build on the existing work on Data model from Red Hat and others in the open source community, demonstrate the work and identify any gaps.
2. To enhance the data model to based on the gaps.

Why this project?

Kubernetes is an open-source system for automating deployment, scaling, and management of containerized applications. It groups containers making up logical units for easy management and discovery. Kubernetes is hosted by the Cloud Native Computing Foundation (CNCF).
Kubernetes builds upon a decade and a half of experience at Google running production workloads at scale using a system called Borg, combined with best-of-breed ideas and practices from the community.
Logging is one of the major challenges with any large deployment on platforms such as Kubernetes, but configuring and maintaining a central repository for log collection can ease the day-to-day operations. For that purpose, the combination of Fluentd, Elasticsearch, and Kibana can create a powerful logging layer on top of Kubernetes clusters. This way the logs can be easily processed, correlated and queried so the troubleshooting becomes much easier and the time to find root causes for problems gets dramatically reduced.

Why me for this project?

Any project needs the passion and perseverance of its participants to succeed. I see myself as a great match for the project. I am a curious, persistent and dependable

person who is eager to broaden his horizons and devote a reasonable amount of time to the work. I am optimistic that the project will benefit from my contributions as much as I will benefit from the project.
I have been contributing and working with the Mozilla Foundation, Docker for over 2 years now in various perspective spanning from community to technical (contributions). With a computer science & engineering background and with a previous year experience in open source projects and hackathons based on Docker and Blockchain, I have the capability to execute this project and determined to learn additional skills that might be needed in the project that I don't yet have at this moment (during the community bonding period). Also, with my programming experience of over 3 years and working with AWS, Bluemix, Blockchain and other related projects, I have the necessary skills it takes to execute this project in 3 months.
My previous Hackathon project was a toolchain based on Hyperledger Fabric developed at Rajasthan Hackathon and this project is employing bash scripting, docker, go on which I will be working with another open source project Content Performance Analytics. With this experience, I am able to complete a GSoC project and these projects uses more than half of the requirements of this project.

Schedule

Milestones

- Collecting logging with Fluentd.
- Setting up fluent-plugin-kubernetes_metadata_input, which allows Kubernetes to create symlinks to Docker log files in **/var/log/containers/*.log**
- Run Fluentd on each of our nodes which is use to launch PodSpec (in YAML/JSON).
- Configure fluentd-pod.yaml to launch the pod.
- Streaming logs from Fluentd into Elasticsearch.
- Querying the logs with Elasticsearch.
- Plugins (these plugins implement the data model).
    - Fluent-plugin-collectd-nest
    - Fluent-plugin-viaq_data_model
    - Docker-kibana
- Analyze Kubernetes logs in Kibana.
- Launch and access Kibana.
    - Show meaningful queries via Kibana that show health of Kubernetes cluster.
    - Show queries that help diagnose a few standard failures in Kubernetes.

- Make a list of gaps in the existing work on Data model.
- Propose enhancements to data model and implement select enhancements.
- Enable cluster level logging using Google Cloud.

Deliverables

- Create pods after creating cluster, the standard output and standard error output for each container is ingested into Elasticsearch and Kibana using a Fluentd agent.
- Configure and launch Elasticsearch as a replication controller.
- Create an Elasticsearch service for communicating with Fluentd.
- Configure and launch Kibana as a replication controller.

Post Google Summer of Code

I am a final year undergraduate of Computer Science currently pursuing Bachelor in B.Tech, and hence have at least two month till I finish my college. I would like to follow up on my project work even after the Google Summer of Code program is over. I will be available for maintenance and bug fixes (if any) in my work even after GSoC, and would very much like to use this experience to work as mentor for upcoming outreach programs.

Timeline

An overview :

Week 1 - Week 4 : Getting to know the community and making a plan including installation, configuration of Fluentd with testing (whether it’s running or not).
Week 4 - Week 8 : Setup the pipeline into Elasticsearch and launch the Elasticsearch replication controller.
Week 8 - Week 12 : Build an Elasticsearch service for communicating with Fluentd and query container logs.
Week 12 - Week 15 : Display Kubernetes logs in Kibana and configure a Kibana service enabling cluster-level logging using Google Cloud.
Week 16 : Wrapping up all subtasks and evaluation of final goal.

Time Availability during GSoC

I would put in 40 hours per week on this project and will be coding during weekends (occasionally), regularly informing my mentors on my progress and regularly updating my Github wiki report page of the project (reporting). I will mostly be programming in the evenings due to college during the day (only May and after that I have my summer break) but if working on Saturdays and/or Sundays, I will probably be programming throughout (day and night).

Profile

Name: Amit Kumar Jaiswal

Email: amitkumarj441@gmail.com

Alternate Email: amit_gkp@live.in

LinkedIn: https://in.linkedin.com/in/amitkumarjaiswal1

GitHub : GitHub Profile Page

Web : http://amitkumarj441.github.io

Skype: amitkumarja

Telegram : @amitkumarjaiswal

Twitter : @AMIT_GKP

CV : https://goo.gl/xmwLPw

Telephone: +91-8081187743

Country of Residence: India

Timezone: India(UTC+05:30)

Primary Language: English

Why CNCF?

Though I was hugely delighted by the experience of open source contribution to Mozilla that I have gained so far and would love to reveal more things to help me understand more about the subject I truly adore, to help me understand the mind of consumer better and to learn some soft skills which have proven to be effective over many years. I would specifically like to focus on projects involving coding, cloud computing and scripting. This will go a long way to improve my community and India including the world as a whole in terms of education and foundations to cloud computing. Since the Indian community is somewhat lacking in the global movement, this is the reason why I decided to forge into the movement, hence filling the gaps.

Open Source Development Experience

My Contributions (technical)

1. Transcedent-AI-Labs(DynaML) [Mozilla Science Lab project] : https://github.com/transcendent-ai-labs/DynaML/commits?author=amitkumarj441
2. Content Performance Analytics [Digital Ocean Hackathon] : https://github.com/ahilles107/content-performance-analytics/commits?author=amitkumarj441
3. STEMMRoleModels [Mozilla Science Lab project] : https://github.com/KirstieJane/STEMMRoleModels/commits?author=amitkumarj441
4. Everware [CERN & Mozilla Science Lab project] : https://github.com/everware/everware/commits?author=amitkumarj441
5. NASA Space Apps Hackathon 2016 : https://github.com/spaceappsuy/olympus-project/commits?author=amitkumarj441
6. Syntax Highlighting [KDE] : https://github.com/KDE/syntax-highlighting/commits?author=amitkumarj441
7. Docker : https://github.com/docker/docker/issues/25981

As mentioned above, Currently, I am coaching few teams of female developers hoping to participate in the RGSoC program in Wikimedia, Scikit-learn and Servo.

About Me

- I'm an active member and and an organizer for several communities like Docker Lucknow, Rust India User Group, Mozilla Science Lab study group, GeeksforGeeks, Brave Software, WikiMedia in our community(India). I have participated in various development meetups organised in our community to sensitise and mentor young, talented and motivated students to contribute to Open Source movements.
- Mozilla Uttar Pradesh is a community that was founded in our state to teach, enhance and empower programming skills to interested students, and I am an admin and mentor in the community and it's Facebook page.
- **Education :** I am currently in my fourth year of undergraduate study at UIET CSJM University, pursuing a four year program which will award me with a

B.Tech in Computer Science & Engineering, with the expected year of graduation being 2017.

My programming skills and qualifications

- Technical Skills
    - Programming Languages : C/C++ (Intermediate), Python (Intermediate), Scala (Intermediate), Go (Intermediate), Ruby (Beginner), Bash (Intermediate)
    - Software Tools : Git/Github/BitBucket, Gerrit, Docker, Ansible, Kubernetes, OpenStack, Elasticsearch
    - Frameworks : Eclipse, JetBrains IDE(except C# IDEs), SublimeText, Feather.js
- Competitive Programming
    - Hackerrank : https://www.hackerrank.com/amit_gkp
    - Hackerearth : https://www.hackerearth.com/@amitkumarj441
    - Kaggle : https://www.kaggle.com/amitkumarjaiswal