

Team 14 Report on Effort Expended PA4

While working on this PA, we came across multiple issues and hangups, however with enough troubleshooting we were able to complete it successfully.

The biggest challenge we kept facing was disk pressure errors on our 4 VM cluster as building all of the docker images that had heavy pip installs caused huge issues. Builds were failing due to disk pressure, and likewise so were our pod deployments. There was no surefire solution to this issue for us so we had to constantly get creative with how we could both lighten our images by using minimized installs, but also deleting unnecessary data from our VMs. This obstacle heavily impeded the speed at which we could complete this project as anytime we had an issue and needed to rebuild an image or redeploy a pod, we would constantly have to clear caches and eke out as much space as possible. This issue highlights the importance of understanding the system you plan to build, and planning ahead in terms of system requirements like memory, space, and power. Overall getting spark to work locally took about 4-5 hours.

Our next step was moving the setup from our 4 VMs to our shared cluster. While we were initially worried that slight differences in our local set up vs on the new cluster would cause a lot of issues, this step was surprisingly clean. For the most part aside from some minor troubleshooting we did not have much trouble deploying our set up on the new clusters. This step took roughly 2-3 hours.

In retrospect, while there was a steep learning curve, this project offered great experience in managing and implementing complex, distributed cloud-based systems. With enough time and research overcoming these obstacles was completely doable and very educational.