IRIS HEP student project

Duration - 12 weeks

Project Student - Kyrylo Meliushko

The CMS monitoring infrastructure is based on five Kubernetes (k8s) clusters. With its growth, more and more open-source project management options have been getting integrated. As for now, lots of cron jobs are running in various environments, writing outputs to different storages, which creates a dependency to the older infrastructure, also including Spark Analytix Cluster managed by CERN IT and ElasticSearch/EOS/MongoDB to store the results. Inter-dependencies may be caused by these cron jobs running Spark jobs, Sqoop jobs, and some Python/Go tasks. The possibility of long term issues of this current design could be the reason to move each cron job to Kubernetes, which would make the development and deployment cycle more efficient and robust, also allowing to orchestrate them easier and safer. The system should be able to flawlessly run complex cron jobs in K8s, provide alerting, logging and monitoring, and have a trusted alert mechanism and a monitoring dashboard to track the cron jobs. Thus, one of the main objectives of this whole project would be to write individual tests for each of the cron jobs to make sure they will get translated to Kubernetes correctly, not causing any failures. The tests would also be useful in deployment and maintenance via alerts. Each cron job has its own unique test script to define its success.

In the last phase of the project, I would like to explore the automation in the deployment of these kubernetes based services using CI/CD pipelines.

Timeline of the Project	
Week:	
1-3	Explore and study the structure of the project, learn bash scripting python libraries like PySpark/Pandas, query languages of ES/MongoDB Service Deployment and management.
4-6	Learn storage operations and other tools of the infrastructure. Learn to deploy test kubernetes clusters with services in development mode
6-9	Analyze and study the performance of services in development mode and then moving the existing cron services on the production kubernetes clusters and commissioning the services in production. Performing sanity tests and testing recovery and failover mechanisms.
9-12	Preparing the project report and exploring automation in deployment using gitlab CI/CD and kubernetes ecosystem.