

Containerized Analytics Continuous processing workflow

Kubernetes and Docker: enabling an efficient, reliable and continuous delivery, with significant increased capacity for running analytics.

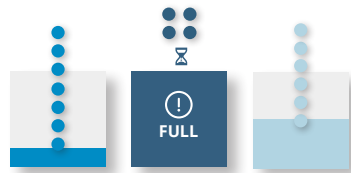
In the context of moving a custom analytics platform to managed, containerized components, we designed and implemented a modern solution to provide reproducible, on-demand analytics batch jobs in a resource efficient manner.

CHALLENGES

Same infrastructure for long-running applications, composed of several simulation, modeling and reporting steps, and on-demand transient analytics workloads.

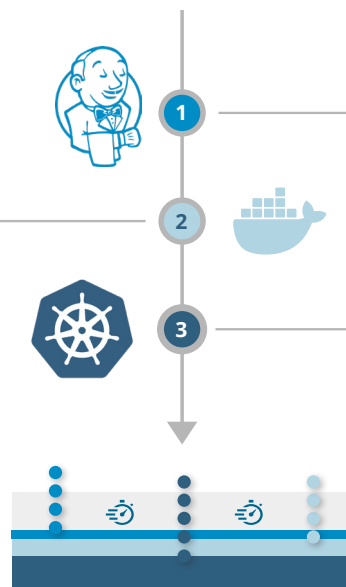
Ensuring **numerical reproducibility** across development, CI and production environments.

Providing support for heterogeneous analytics workloads with run-times varying between seconds and multiple hours.



SOLUTIONS

Configure **containerized development environments** (RStudio sessions running on Kubernetes).



Continuous integration builds with **Jenkins** producing container images based on versioned code packages.

Replaced commercial solution for submitting job requests to a dedicated cluster with a custom component for managing analytic job requests on Kubernetes.



Numerical reproducibility: Using container images enables better control over underlying dependencies and specific configurations, ensuring that calculations are reproducible across all environments.



Flexible resource allocation: Kubernetes allows pooling resources across multiple machines between all components. All free resources can be used by analytic jobs, instead of only a dedicated cluster.



“ We were quickly able to run a prototype in parallel with the legacy solution and validate our approach, which gave us the needed confidence to build the complete solution.