Xavier Espinal, 6 May 2020

Proposal: ESCAPE Data Lake Deployment and Operations

- We are reaching a reasonable scale in our ESCAPE Data Lake, the deployment and operations load start to be sensible

- Time to set-up a group of people that coordinate deployment and follow-up operations in a distributed fashion

- Short term goal is to be **effective in building up a prototype** at a scale which is representative for the sites and the sciences.

  - We need to be able to do a **performance** evaluation and also have a realistic **operations** *taste*.

- The goal long term is fundamental, and is to ensure **sustainability** of the ESCAPE model and the tools after the end of the project.

  - Have people at the sites and involved in the experiments that have experience in a **distributed computing environment**.
  - Have people at the sites and involved in the experiments that have experience deploying and operating some of the tools we are using: RUCIO, FTS, CRIC, AAI, etc.
  - Hence the eventual adoption of the Data Lake concept and the tools within the experiments will be more likely and much easier.

- Operating and helping to the deployment of an early model provides unique opportunity to **build-up expertise** (c.f. early operations and deployment teams in LCG ~2005)

- The caveat is manpower and resources. The idea is to frame this as a best-effort activity. This should be also considered as an investment to the site and to the experiment, and of course to enlarge the expertise and technical knowledge of the designated persons that will be doing the work. It might look like the opposite, but this will be fun (I know by experience).

  - What about a team of ~8 people to start with?
    - e.g. if we say we dedicate 2 hours per day overall to see how the infrastructure is working, this means 16 hours per week in total which translates into 2 hours/person/week (and some coordination chat/mails)

- Reference Tools:

  - **Monitoring** is being set up: dashboard providing live info for networking (perfsonar tests), automated transfer machinery (file put/get/delete), smoke tests for file movement and QoS transitions.
  - JIRA **ticketing system** to be provided by SKAO. Allow for centralized information point, knowledge base and tracking.