Total Cost of Ownership and Evaluation of Google Cloud Resources for the ATLAS Experiment at the LHC

Gonzalo Merino[1] and *David* South[2,*]
on behalf of ATLAS Software and Computing

[1]Port d'Informació Científica, Barcelona, Spain
[2]Deutsches Elektronen–Synchrotron (DESY), Hamburg, Germany

Abstract. The ATLAS Google Project was established as part of an ongoing evaluation of the use of commercial clouds by the ATLAS Collaboration, in anticipation of the potential future adoption of such resources by WLCG grid sites to fulfil or complement their computing pledges. Seamless integration of Google cloud resources into the worldwide ATLAS distributed computing infrastructure was achieved at large scale and for an extended period of time, and hence cloud resources are shown to be an effective mechanism to provide additional, flexible computing capacity to ATLAS. For the first time a Total Cost of Ownership analysis has been performed, to identify the dominant cost drivers and explore effective mechanisms for cost control. Network usage significantly impacts the costs of certain ATLAS workflows, underscoring the importance of implementing such mechanisms. Resource bursting has been successfully demonstrated, whilst exposing the true cost of this type of activity. A follow–up to the project is underway to investigate methods for improving the integration of cloud resources in data–intensive distributed computing environments and reducing costs related to network connectivity, which represents the primary expense when extensively utilising cloud resources.

ATL-SOFT-PROC-2025-037
24 February 2025

As the type of computing resources adopted by WLCG sites to either fulfil or complement their computing pledges continues to diversify, the ATLAS Google Project was established to evalute the suitability of commercial cloud for this purpose. This involved running a cloud site in production for an extended period of time, from July 2022 until September 2023, in order to gain operational experience at a large scale. An evaluation of all ATLAS workflows was performed, including all Monte Carlo production steps, data reprocessing, and user analysis, as well as demonstrating rapid and efficient bursting to additional, large scale resources. An important aspect to this project was the appraisal of a new Google Cloud Platform fixed–price subscription model, in order to evaluate the Total Cost of Ownership of employing a commercial cloud site at scale. Further R&D work, building upon experience already gained, was also done to exploit unique opportunities available at Google, such as GPUs and ARM CPUs. These proceedings are intended to be a brief summary of this conference contribution; a full description of this work may be found in Ref. [1].

Commercial cloud providers employ highly detailed, granular and itemised cost breakdowns for all services used during a contract, based on the so called *list–price*. Conversely, evaluating the effective costs of the usual resources employed by ATLAS is often difficult

*e-mail: david.south@desy.de

due to a large variation in funding methodologies with respect to the resource type, country, and even from different funding agencies within the same country. The situation is also made difficult via unknown local arrangements, for example where some sites, typically at universities, do not directly pay electricity or WAN access costs. Therefore, it is important to point out that the TCO evaluation performed does not attempt to directly compare the cost of running ATLAS jobs in the Google Cloud with the cost of running an ATLAS grid site.

A *Service Agreement* contract was negotiated with Google, based on resources usage estimates, which features a significant discount with respect to the list–price. The agreement had a flat rate cost of around $57k/month with CPU cores provided as *spot instances*, meaning pre–emption is possible. Taking this literally, the TCO is the total price over the duration of the 15 month contract, around $850k, although this number is not really informative or useful. Therefore, this evaluation has examined the relative contributions of compute, storage and network to the total list–price cost under different operating models and for different workflows, to understand how to best use such a resource.

The project successfully demonstrated that deploying an effective ATLAS site in a commercial cloud at a very large scale requires little operational effort. The current ATLAS workflow and data management tools proved to be adequate to adjust the cloud site configuration, and the solutions developed are essentially are cloud–agnostic, avoiding vendor lock–in.

Commercial cloud computing is found to be an ideal solution for providing additional, stable CPU resources, with an observed overall rate of 5% lost wall–clock from failed jobs over the duration of the project (20% of which came from pre-emptions), resulting in an observed eviction rate between 1-2%, which compares favourably to more traditional grid sites. The CPU provided by Google was homogeneous throughout the project and the PanDA queue associated with the ATLAS Google site experienced negligible downtime. Whilst it involved significantly more work, the seamless integration of cloud–based storage was also achieved, although the network costs incurred from egress may be significant. Some ATLAS workflows are found to be better than others with respect to egress, where data reprocessing being the heaviest, and ATLAS cannot yet say commercial cloud is suitable for all ATLAS workflows. Resource bursting was been shown to be very effective, albeit at significant cost with respect to the list price.

The TCO evaluation showed that without a subscription agreement, the cost of commercial cloud resources is significantly greater and that a favourable agreement that makes sense to both the cloud provider and the client is clearly a necessity. The Google Cloud resources used during this project cost a total of $3.2M at list–price compared to the $850k paid via the subscription agreement, representing a discount of 73%. In other words, the resources used during this project would have been on average 270% more expensive at list–price. This is most obvious in the costs associated with the bursting tests, which at $25k each would have incurred a daily expenditure considerably in excess of the $1900/day rate of the subscription agreement.

By leveraging the Google Cloud Subscription Agreement pricing model, ATLAS has effectively harnessed between three to four times the resources compared to what the same investment would deliver for the list–price. The subscription pricing model applied in this project has proven to be very suitable to ATLAS, although the same rate is not guaranteed going forward and any subsequent contract and any future costs incurred may be higher.

References

[1] ATLAS Collaboration, "Total Cost of Ownership and Evaluation of Google Cloud Resources for the ATLAS Experiment at the LHC," Comput. Softw. Big Sci. **9** (2025) no.2, https://doi.org/10.1007/s41781-024-00128-x,[arXiv:2405.13695 [cs.DC]]