LHC@home
Volunteer computing for the LHC

A World-wide DataBridge Supported By A Commercial Cloud Provider

File Transfer Challenges In Volunteer Computing

One of the core challenges in Volunteer Computing (VC) is how to reliably collect large output files from the volunteers' machines. Due to the highly distributed nature of VC infrastructure, volunteers can be thousands of miles away from the VC project's storage location. Larger distances between the endpoints results in a higher RTTs and an increased chance of packet loss during transmission, thus lowering the transfer performance.

Effect Of RTT On File Transfer Over HTTP

To measure the effect of RTT on file transfers, seven identical virtual machines (VMs) were provisioned at different global locations. An Amazon Web Service (AWS) S3 bucket in the Frankfurt region was used as the destination. Each of the instances performed sequential file transfers over HTTP with file size from 50MB to 300MB in 50MB increments. The results are shown in Figure 1, which compares the average transfer time to the average RTT.

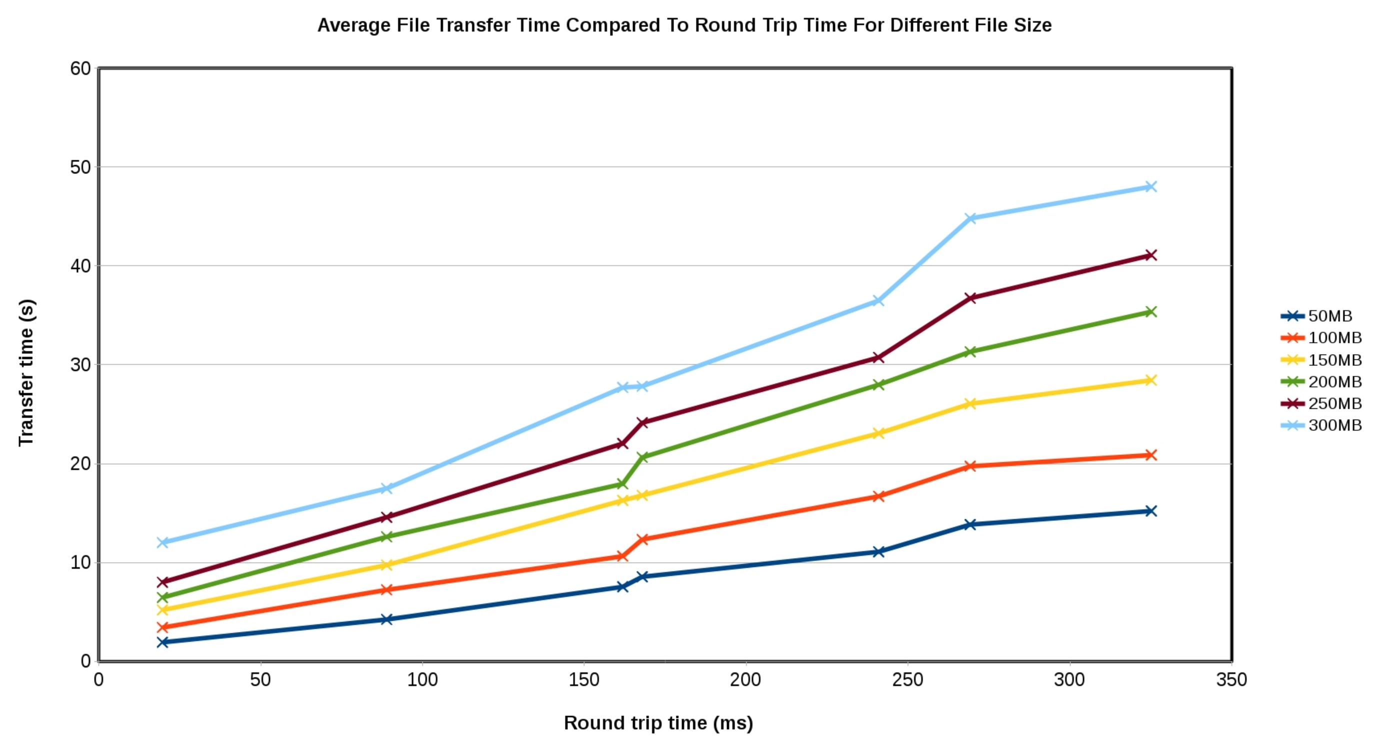

Figure 1. Average file transfer time compared to RTT

It is evident that even for the same file size, transfers from instances where a higher latency was recorded took a considerably longer time to complete. This issue is further compounded by the asymmetrical bandwidth offered by ISPs in residential networks, where upload bandwidth is often many times lower than download bandwidth, as well as poor network infrastructure in some areas. The combined effect of these factors are challengeing for VC projects that produce large outputs. One such project is CMS@home, which runs event simulations for the CMS experiment based at CERN. We analysed the effect of large distances on file transfer performance and the stage-out failures experienced in CMS@home, to understand if there is a correlation between distance and file transfer failure.

Stage-out Failures In CMS@home

Stage-out failures in CMS@home recorded during a two week period were analysed. Figure 2 shows the percentage of each status code recorded for upload attempts from volunteers against the CEPH S3 storage instance at CERN, ordered by their distance from CERN.

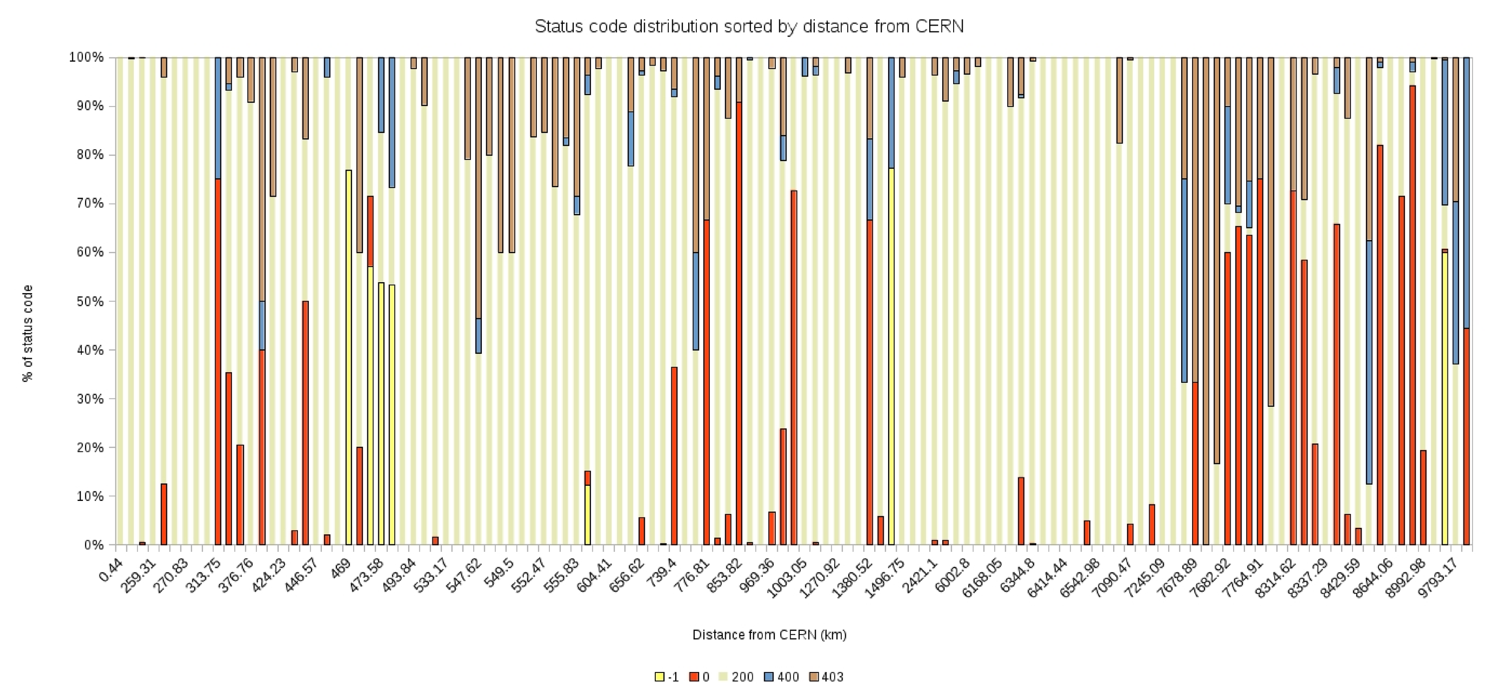

Figure 2. Status code distribution for CMS@home stage-out

It can be seen that the network related error codes (0, 400, 403) occurred more frequently as the distance from CERN increases. This suggests that in a globally distributed VC infrastructure with large output file sizes, the distance between volunteers and the storage endpoint can be a significant factor that affects the success rate of file transfers.

Using AWS S3 To Provide Local Upload Endpoints

The objective of this design is to optimise the global transfer of files by offering the volunteers a closer upload endpoint to an asynchronous eventually consistent method for migrating those files to CERN, or to an alternative data centre where the validation and merge operations can take place. This has the effect of reducing latency between the volunteer's machine and the upload endpoint and hence increases file transfer performance and reliability.

AWS S3 was chosen because of the location of its regions, as shown in Figure 3, which offers a good match with areas identified where there are a high number of volunteers. In practice, storage systems from other providers could also be federated into the system.

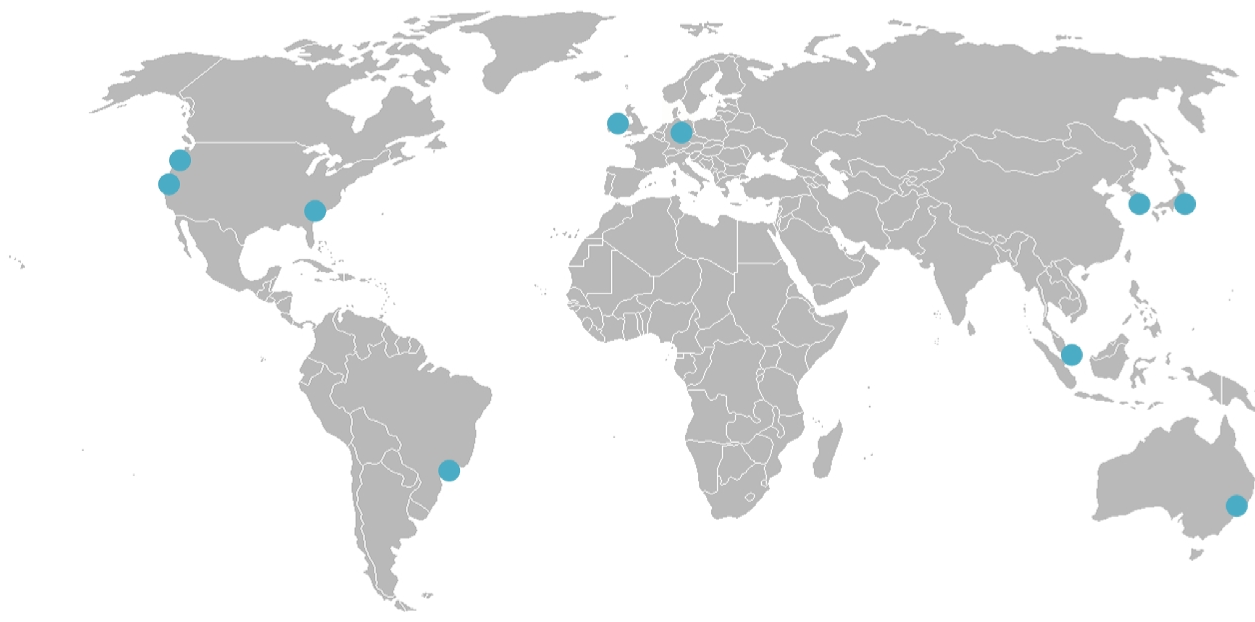

Figure 3. Location of federated upload endpoints

DataBridge Redirection

A core component of the design is the DataBridge, which federates the upload endpoints, and redirects an upload request to an endpoint that is closest to the volunteer. A high-level view of the design is shown in Figure 4, with the flow of events summarised as follows:

1. The volunteer uploads the file via HTTP PUT to the DataBridge.
2. The DataBridge redirects upload requests to the S3 bucket closest to the volunteer.
3. Files uploaded to S3 are replicated in AWS region closest to the project data storage.
4. Files are replicated to the project's data storage from this close data store.

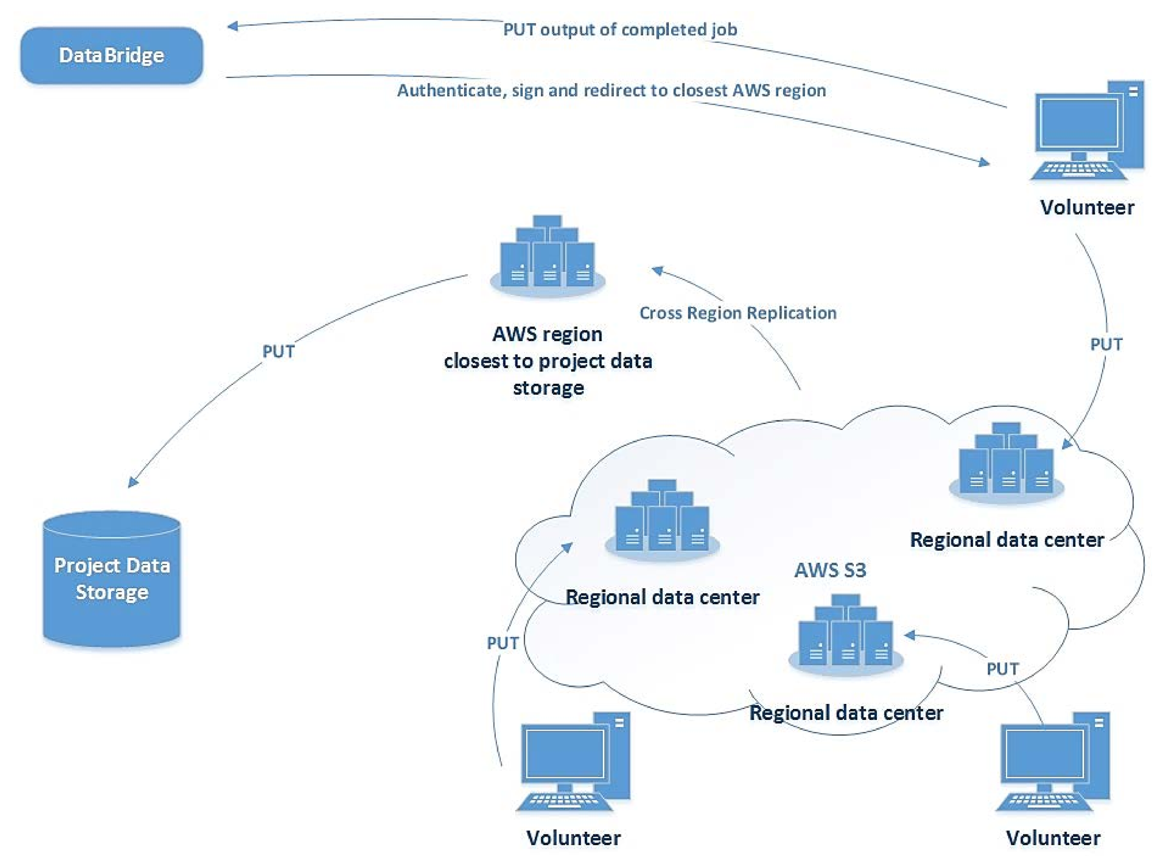

Figure 4. High-level view of a global DataBridge

The advantages of this design is that the latency between the volunteer's machine and the upload endpoint can be reduced, as well as transferring scalability and operational responsibility of the upload endpoints to the commercial provider. The approach is also designed to be highly extensible. As providers expand their global footprint, more locations can be federated into the system.

Evaluating The Design

To evaluate the new design, a DataBridge service was provisioned in Frankfurt and configured to federate ten S3 buckets (one in each AWS region). Another machine was provisioned in Tokyo to act as a volunteer which attempted to transfer files from 50MB to 300MB in 50MB increments over HTTP. In the first test, the transfers were targeted directly at an S3 bucket in the Frankfurt region. In the second, transfers were targeted at the DataBridge, with the requests redirected towards an S3 bucket closest to Tokyo (in this case, a bucket in the Seoul region). The result is shown in Figure 5, which compared the average file transfer time to file size.

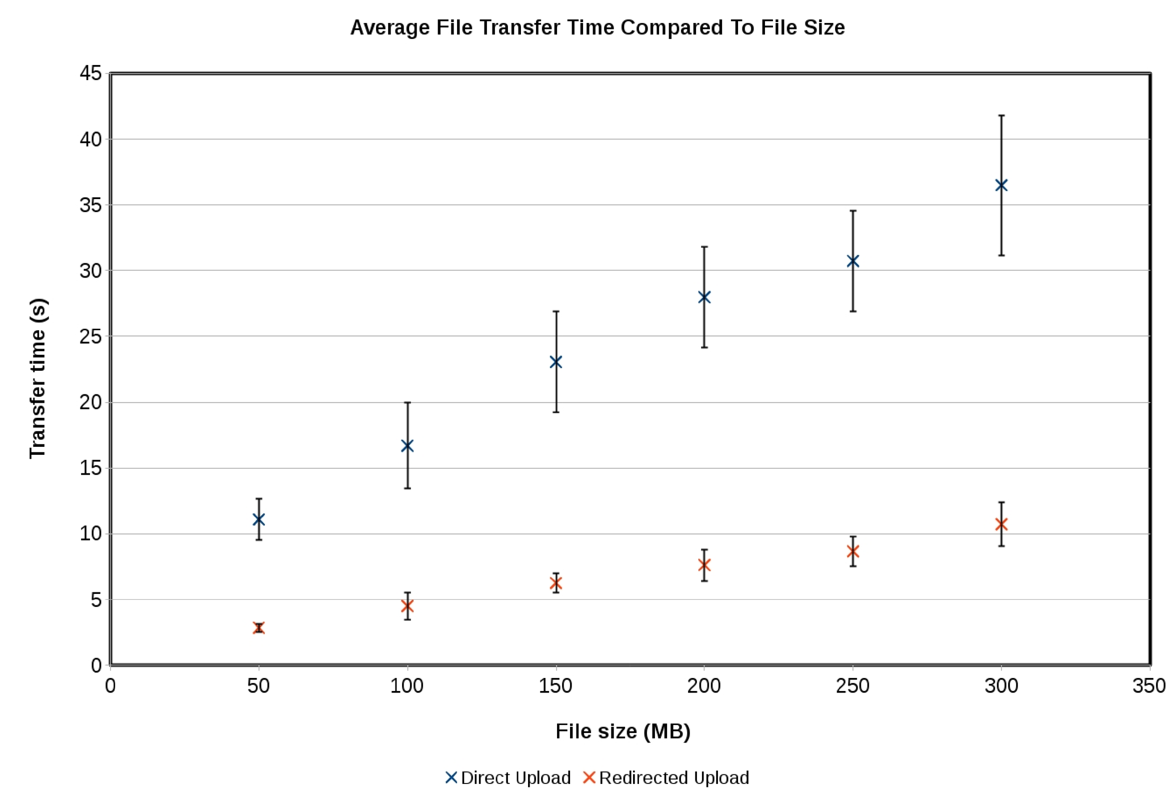

Figure 5. Average file transfer time for direct upload and DataBridge redirected upload

The result shows a significant improvement in transfer time when the transfers were redirected towards an endpoint closer to the Tokyo instance. Further tests also shows that file uploads using this approach are more resilient to performance penalty caused by poor network conditions, when a 2% packet loss was introduced.

Kwong Tat Cheung - University of the West of England
Laurence Field - CERN
Fabrizio Furano - CERN

CERN