Contribution ID: **14**

Type: **Presentation**

Rocket to the Cloud – A Faster Way to Upload

Monday, January 29, 2018 2:00 PM (20 minutes)

Rocket is the first attempt at handling one of the particular problems that other tools have failed to solve. This presentation will demonstrate AARNet's experiences and tools used high-speed data transfers of different kinds of research data.

The research community in Australia is spread far and wide geographically, resulting in some cases to be physically far from one of our three CloudStor sites spread across the country. In addition, the data sets researchers store can be very varied, ranging from ephemeral data, to archival data. From many small files, to fewer very large files. This has meant that AARNet's software infrastructure needs to be spread in order to minimise network latencies between nodes, and this has created its own challenges in providing a reliable and reusable platforms for data sharing and transfer. Managing these requirements has resulted in more than one way to upload and share data.

Rocket helps some users run scientific instrumentation and require a tool to quickly upload vast amounts of datasets quickly. Usually the ownCloud sync client is used but for some it is not quick enough because it uploads files one by one with a single thread via the ownCloud webdav gateway, which can choke when presented many little files. In addition, the sync client is geared more to synchronisation rather than just upload, meaning that both client and server store the same data. This is undesired by instrument users as it causes issues and interrupts the natural workflow. In some cases the instrument PC's disk becomes full of synchronised data they do not need. For this reason, we have developed a product called Rocket, which integrates directly into ownCloud and EOS.

Rocket is an upload only tool that bundles and uploads payloads of data into CloudStor using parallel threads. A payload can consist of bundles small files and chunks of large files. Settings such as payload sizes, number of threads, maximum number of files per payload, payloads to buffer in memory is all user modifiable. This means users can fine tune settings to best utilise their local network and PCs.

By keeping payload sizes consistent and by using parallel threads, in the right conditions, we are able to upload data as fast as the PC can read off the local disk. Rocket uploads files into our ownCloud, so it is possible to upload into a shared space and have files arrive at a group of users as files are uploaded.

Primary author: D'SILVA, Michael (AARNet)

Co-author: JERICHO, David (AARNet)

Presenter: D'SILVA, Michael (AARNet)

Session Classification: Technology: deployment, testing and optimization