



Towards Cloud Processing of GGOS Big Data

Stuart Weston (1), Bumjun Kim (1), Alan Litchfield (1), Sergei Gulyaev (1), Dylan Hall (2), Carlos Choraó (2), Andrew Ruthven (3), Glyn Davies (3), Bruno Lagos (3), and Don Christie (3)

(1) Institute for Radio Astronomy & Space Research, AUT University, Auckland, New Zealand (sweston@aut.ac.nz), (2) Research and Education Advanced Network New Zealand, Wellington, New Zealand, (3) Catalyst IT Ltd, Wellington, New Zealand

We report on our initial steps towards development of a cloud-like correlation infrastructure for geodetic Very Long Baseline Interferometry (VLBI), which in its raw format is of the order of 10-100 TB (big data). Data is generated by multiple VLBI radio telescopes, and is then used by for geodetic, geophysical, and astrometric research and operational activities through the International VLBI Service (IVS), as well as for corrections of GPS satellite orbits. Currently IVS data is correlated in several international Correlators (Correlation Centres), which receive data from individual radio telescope stations either in hard drives via regular mail service or via fibre using e-transfer mode. The latter is strongly limited by connectivity of existing correlation centres, which creates bottle necks and slows down the turnover of the data. This becomes critical in many applications – for example, it currently takes 1-2 weeks to generate the dUT1 parameter for corrections of GNSS orbits while less than 1-2 days delay is desirable. We started with a blade server at the AUT campus to emulate a cloud server using Virtual Machines (VMWare). The New Zealand Data Head node is connected to the high speed (100 Gbps) network ring circuit courtesy of the Research and Education Advanced Network New Zealand (REANNZ), with the additional nodes at remote physical sites connected via 10 Gbps fibre. We use real Australian Long Baseline Array (LBA) observational data from 6 radio telescopes in Australia, South Africa and New Zealand (15 baselines) of 1.5 hours in duration making 8 TB to emulate data transfer from remote locations and to provide a meaningful benchmark dataset for correlation. Data was successfully transferred using bespoke UDT network transfer tools and correlated with the speed-up factor of 0.8 using DiFX software correlator. In partnership with the New Zealand office of Catalyst IT Ltd we have moved this environment into Catalyst Cloud and report on the first correlation of a VLBI Dataset in a true cloud environment.