The Prototype of an Object Storage for Distributed Acoustic Sensing

Yiyu Ni[1]*, Marine A. Denolle[1], Naomi Alterman[2], Rob Fatland[2], Chengxin Jiang[4], Bradley P. Lipovsky[1], Fredrich Knuth[3]
1) Department of Earth and Space Sciences, University of Washington
2) eScience Institute, University of Washington
3) Department of Civil and Environmental Engineering, University of Washington
4) Research School of Earth Sciences, the Australian National University
*niyiyu@uw.edu

Distributed Acoustic Sensing (DAS) is a new seismic observation method. It utilizes repeated laser pulses along optical fibers up to 100 km in length to measure phase changes of backscattered light that occur due to rapid straining rate of the fiber. DAS dramatically expands the capability of dense seismic observation and has been used for buried tectonic fault detection and near-surface imaging. However, the huge amount of data generated by the DAS challenges the data I/O from both local computing clusters and data centers perspective, hence limiting the processing required for seismological research in general. Commercial cloud environments are promising computing architecture for such large-scale data processing, yet the current and standard DAS data formats are not optimized for cloud-native object storage.

In this study, we propose a data platform for DAS data that deploys an object storage with cloud-optimized data formats. This high-performance framework is able to host data for users in individual research groups or institutions that own mid-scale Linux servers. The platform also uses the same API with AWS Cloud storage (Simple Storage Service, S3). Allowed by this S3-compatibility, we migrate this framework and host one month of urban DAS data on AWS S3. An ambient noise cross-correlation workflow is implemented with data I/O fully through our platform using 1 month of data and 600 channels. The hourly cross-correlation function of 180k channel-pairs is calculated and saved to S3 on the fly. With AWS batch service and auto-scaling, 8 billion cross-correlation functions are computed in parallel and finished in 4 hours with 67 virtual machines of 4 CPUs and 6GB RAM. The total cost of the cross-correlation workflow is less than $11. This study shows that, while natively supports data and product sharing, our data platform is capable of serving massive DAS cross-correlation.