NAME: ENOCH DADZIE
COURSE: GE562
INDEX NUMBER: PG2896920

Based on Gorelick et al (2017), briefly describe the architecture of the Google Earth Engine (GEE). Discuss the advantages and limitations of GEE compared to convential remote sensing

Google Earth Engine combines a multi-petabyte catalog of satellite imagery and geospatial datasets with planetary-scale analysis capabilities. Scientists, researchers, and developers use Earth Engine to detect changes, map trends, and quantify differences on the Earth's surface.

The architecture of the google earth engine is built in a way that the earth engine code editor and third-party applications use client libraries to send interactive or batch queries to the system through a REST API. On-the-fly requests are handled by Front End servers that forward complex sub-queries to Compute Masters, which manage computation distribution among a pool of Compute Servers. The batch system operates in a similar manner, but uses flumeJava to manage distribution. Backing both computation systems are a collection of data services, including an asset database that contains the per-image metadata and provides efficient filtering capabilities. The borg cluster management software manages each component of the system and each service is load-balanced over multiple workers. Failure of any individual worker just results in the caller reissuing the query.

Earth Engine is used across a wide variety of disciplines, covering topics such as global forest change, global surface water change, crop yield estimation, rice paddy mapping, urban mapping, flood mapping, fire recovery and malaria risk mapping. It has also been integrated into a number of third-party applications, for example analyzing species habitat ranges, monitoring climate, and assessing land use change. The details of a few of these applications illustrate how Earth Engine's capabilities are being leveraged compared to conventional remote sensing which mostly specialize in a specific discipline.

Google Earth Engine create an interface that can display time-series mapping information, such as from the Moderate Resolution Imaging Spectroradiometer (MODIS) imagery, with climate data to enable scientists to more easily obtain surface climate-based data relevant for earth observation. This saves time and enables tasks that were previously the privy of programmers to be applied by scientists directly.

Benefits of the Google Earth Engine include the integration of different datasets at different resolution, in this case Sentinel-1, Sentinel-2, and Landsat 8 imagery, over different time intervals such that gaps in data and improvement of identification of features is possible through multi-image integration. The result helps researchers to apply temporal aggregation of data to fill in major gaps due to cloud coverage that normally would have involved downloading many images and then processing each for integration before then conducting identification analyses that classify land cover features.

Earth Engine is designed to support fast, interactive exploration and analysis of spatial data, allowing the user to pan and zoom through results to examine a subset of the image at a time. To facilitate this, Earth Engine uses a lazy computation model that allows it to compute only the portions of output that are necessary to fulfill the current request. For instance, a user might wish to compute the difference between two seasonal composites, to highlight the changes due to phenology or snow-cover.

Many raster processing operations used in remote sensing are local: the computation of any particular output pixel depends only on input pixels within some fixed distance. Example include per-pixel operations such as band math or spectral unmixing, as well as neighborhood operations such as convolution or texture analysis. Using earth engine, these operations can be easily processed in parallel by subdividing an area into tiles and computing each independently. Processing each output tile usually requires retrieving only one or a small number of tiles for each input. This fact, combined with pyramided inputs and judicious caching, allows for fast computation of results at any requested scale or projection.

The system of google earth engine is entirely responsible for deciding how to run a computation. This results in some interesting challenges in both the design and use of the system. While this model enables massive parallelism in distributed commodity servers, it does so at the expense of introducing a complex coding style. As it has been noticed in the architecture of the google earth engine the combination of server and client-side programming tends to be confusing. For example, simple index iteration is not recommended when working with server-side objects, because the index itself is a client-side variable. To iterate an image collection, we must define a certain recursive function, which cannot modify values outside of the function’s scope, among other limitations. By this in mind, it is quite clear that porting sophisticated applications into the GEE framework can be quite challenging

To use Google Earth Engine, compliance with the license agreement is required. The license agreement states explicitly Earth Engine's terms allow for use in development, research, and education environments. It may also be used for evaluation in a commercial or operational environment, but sustained production use is not allowed. Additionally, data products generated by Earth Engine may not be sold thus making it free for non-commercial use. In sum, Google Earth Engine is targeted at non-profit institutions. And even though a good number of research is carried out in a non-for-profit spirit, many researchers could be interested in potential bonuses of their research

Also, the free version is not suited for production workloads. GEE's FAQ also states that sustained production use is not allowed. So even if we are working in a non-profit setting, we can't rely on GEE for goals such as continued, real-time environmental monitoring. Sustained production use is not only in violation of GEE's terms and services, it is also impossible to do, as workloads with estimated times to completion longer than 5 minutes will be executed in batch mode, at indeterminate times. Presumably, if a user is in violation of the terms and services, GEE will stop executing those batch jobs.

The free version has some processing and storage limits, While GEE is generally free, there are a few caveats concerning processing time and storage limits. As for processing, there are two available modes within GEE: interactive and batch-mode. The interactive mode is extremely fast but limited to jobs that can be served in under 5 minutes of processing time. The batch mode, on the other hand, is considerably slower and requires the export of data to various Google's services, such as Cloud Compute or Google Drive, where users can incur some costs, albeit probably small ones.

GEE is inconvenient for smaller datasets. While not exactly a limitation of GEE, something else to take into consideration is that GEE might not be necessary at all in many cases, as many important remote sensing datasets can be handled perfectly by desktop computers. And a system that is designed for handling massive parallelism and Petabyte-scale datasets isn't the optimal choice for dealing with smaller datasets. A good number of remote sensing applications require fast access to data that can perfectly fit in just a few Terabytes, or even a few hundred GB. SSD drives currently sell for under $40 per TB and even an inexpensive laptop is extremely powerful nowadays. Additionally, open-source software libraries, which is in some cases are provided by Space Agencies themselves, and importantly, under permissive copyleft licensing conditions, are making great progress in easing the very first few steps of the Remote Sensing pipeline (downloading, opening, and creating time-series of data).

In sum, GEE should be considered as a rapid-prototyping tool for Geospatial applications.

In cases where the possibility of getting involved in commercial applications or sustained production use of remote sensing, taking the time to acquire the skills to work independently from Google Earth Engine can be worthwhile, especially as the cost of powerful desktop workstations and hard drives are going down, while the quality of open-source remote sensing software licensed under permissive conditions is going up.