ARTICLE IN PRESS
RSE-10315; No of Pages 10
Remote Sensing of Environment xxx (2016) xxx–xxx
Contents lists available at ScienceDirect
Remote Sensing of Environment
journal homepage: www.elsevier.com/locate/rse

Google Earth Engine: Planetary-scale geospatial analysis for everyone

Noel Gorelick [a,*], Matt Hancher [b], Mike Dixon [b], Simon Ilyushchenko [b], David Thau [b], Rebecca Moore [b]

[a] Google Switzerland, Brandschenkestrasse 110, Zurich 8002, Switzerland
[b] Google Inc., 1600 Amphitheater Parkway, Mountain View, CA, 94043, USA

ARTICLE INFO

Article history:
Received 9 July 2016
Received in revised form 5 June 2017
Accepted 27 June 2017
Available online xxxx

Keywords:
Cloud computing
Big data
Analysis
Platform
Data democratization
Earth Engine

ABSTRACT

Google Earth Engine is a cloud-based platform for planetary-scale geospatial analysis that brings Google's massive computational capabilities to bear on a variety of high-impact societal issues including deforestation, drought, disaster, disease, food security, water management, climate monitoring and environmental protection. It is unique in the field as an integrated platform designed to empower not only traditional remote sensing scientists, but also a much wider audience that lacks the technical capacity needed to utilize traditional supercomputers or large-scale commodity cloud computing resources.

1. Introduction

Supercomputers and high-performance computing systems are becoming abundant (Cossu et al., 2010; Nemani et al., 2011) and large-scale cloud computing is universally available as a commodity. At the same time, petabyte-scale archives of remote sensing data have become freely available from multiple U.S. Government agencies including NASA, the U.S. Geological Survey, and NOAA (Woodcock et al., 2008; Loveland and Dwyer, 2012; Nemani et al., 2011), as well as the European Space Agency (Copernicus Data Access Policy, 2016), and a wide variety of tools have been developed to facilitate large-scale processing of geospatial data, including TerraLib (Câmara et al., 2000), Hadoop (Whitman et al., 2014), GeoSpark (Yu et al., 2015), and GeoMesa (Hughes et al., 2015).

Unfortunately, taking full advantage of these resources still requires considerable technical expertise and effort. One major hurdle is in basic information technology (IT) management: data acquisition and storage; parsing obscure file formats; managing databases, machine allocations, jobs and job queues, CPUs, GPUs, and networking; and using any of the multitudes of geospatial data processing frameworks.

This burden can put these tools out of the reach of many researchers and operational users, restricting access to the information contained within many large remote-sensing datasets to remote-sensing experts with special access to high-performance computing resources.

Google Earth Engine is a cloud-based platform that makes it easy to access high-performance computing resources for processing very large geospatial datasets, without having to suffer the IT pains currently surrounding either. Additionally, and unlike most supercomputing centers, Earth Engine is also designed to help researchers easily disseminate their results to other researchers, policy makers, NGOs, field workers, and even the general public. Once an algorithm has been developed on Earth Engine, users can produce systematic data products or deploy interactive applications backed by Earth Engine's resources, without needing to be an expert in application development, web programming or HTML.

2. Platform overview

Earth Engine consists of a multi-petabyte analysis-ready data catalog co-located with a high-performance, intrinsically parallel computation service. It is accessed and controlled through an Internet-accessible application programming interface (API) and an associated web-based interactive development environment (IDE) that enables rapid prototyping and visualization of results.

The data catalog houses a large repository of publicly available geospatial datasets, including observations from a variety of satellite and aerial imaging systems in both optical and non-optical wavelengths, environmental variables, weather and climate forecasts and hindcasts, land cover, topographic and socio-economic datasets. All of this data is preprocessed to a ready-to-use but information-preserving form that allows efficient access and removes many barriers associated with data management.

Users can access and analyze data from the public catalog as well as their own private data using a library of operators provided by the Earth Engine API. These operators are implemented in a large parallel

* Corresponding author.
E-mail address: gorelick@google.com (N. Gorelick).

http://dx.doi.org/10.1016/j.rse.2017.06.031

processing system that automatically subdivides and distributes computations, providing high-throughput analysis capabilities. Users access the API either through a thin client library or through a web-based interactive development environment built on top of that client library (Fig. 1).

Users can sign up for access at the Earth Engine homepage, https://earthengine.google.com, and access the user interface, as well as a user's guide, tutorials, examples, training videos, function reference, and educational curricula. While prior experience with GIS, remote sensing and scripting make it easier to get started, they are not strictly required, and the user's guide is oriented towards domain novices. Accounts come with a quota for uploading personal data and saving intermediate products, and any inputs or results can be downloaded for offline use.

3. The data catalog

The Earth Engine public data catalog is a multi-petabyte curated collection of widely used geospatial datasets. The bulk of the catalog is made up of Earth-observing remote sensing imagery, including the entire Landsat archive as well as complete archives of data from Sentinel-1 and Sentinel-2, but it also includes climate forecasts, land cover data and many other environmental, geophysical and socio-economic datasets (Table 1). The catalog is continuously updated at a rate of nearly 6000 scenes per day from active missions, with a typical latency of about 24 h from scene acquisition time. Users can request the addition of new datasets to the public catalog, or they can upload their own private data *via* a REST interface using either browser-based or command-line tools and share with other users or groups as desired.

Earth Engine uses a simple and highly general data model based on 2D gridded raster bands in a lightweight "image" container. Pixels in an individual band must be homogeneous in data type, resolution and projection. However, images can contain any number of bands and the bands within an image need not have uniform data types or projections. Each image can also have associated key/value metadata containing information such as the location, acquisition time, and the conditions under which the image was collected or processed.

Related images, such as all of the images produced by a single sensor, are grouped together and presented as a "collection". Collections provide fast filtering and sorting capabilities that make it easy for users to search through millions of individual images to select data that meets specific spatial, temporal or other criteria. For example, a user can easily select daytime images from the Landsat 7 sensor that cover any part of Iowa, collected on day-of-year 80 to 104, from the years 2010 to 2012, with less than 70% cloud cover.

Images ingested into Earth Engine are pre-processed to facilitate fast and efficient access. First, images are cut into tiles in the image's original projection and resolution and stored in an efficient and replicated tile database. A tile size of 256 × 256 was chosen as a practical trade-off between loading unneeded data *vs.* the overhead of issuing additional reads. In contrast to conventional "data cube" systems, this data ingestion process is information-preserving: the data are always maintained in their original projection, resolution and bit depth, avoiding the data degradation that would be inherent in resampling all data to a fixed grid that may or may not be appropriate for any particular application.

Additionally, in order to enable fast visualization during algorithm development, a pyramid of reduced-resolution tiles is created for each image and stored in the tile database. Each level of the pyramid is created by downsampling the previous level by a factor of two until the entire image fits into a single tile. When downsampling, continuous-valued bands are typically averaged, while discrete-valued bands, such

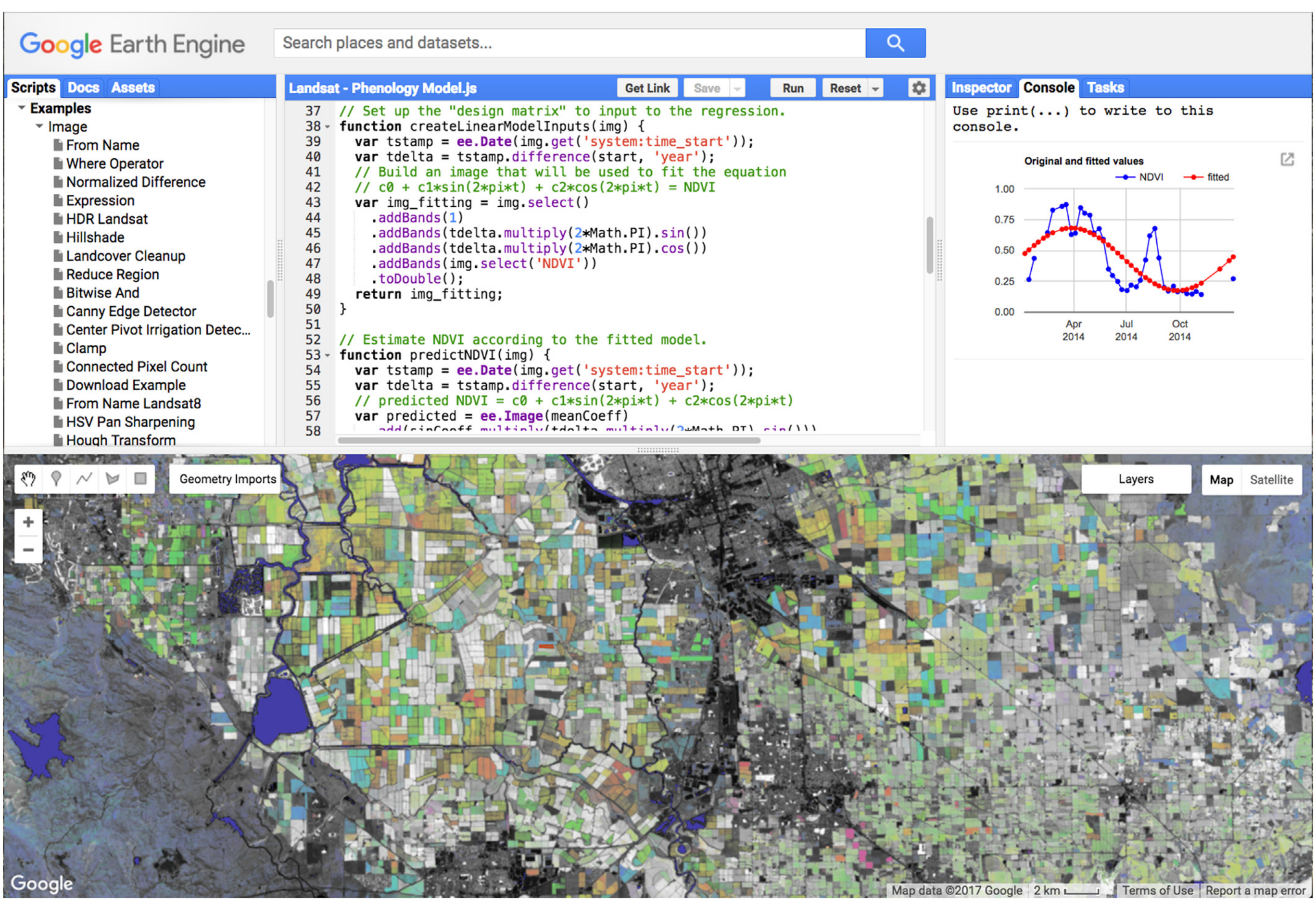

Fig. 1. The Earth Engine interactive development environment.