Aaron Gussman:

Hey, good morning everyone. I'm going to go ahead and share my screen. Oh, let me try that again. All right. All right, can I get a thumbs up when someone can see it? Awesome, thank you. Great.

Hey everyone. My name is Aaron Gussman, I'm a customer engineer with Google Cloud supporting US federal government customers. I have a background in AI and ML in geospatial. I'm really excited to talk to you today. So excited that I literally and figuratively tried to cram the whole Earth into my slide deck here. I have an unprofessional number of slides, so I'm just going to go through them extremely fast, and hopefully leave a little bit of time for questions at the end. The main takeaway that I want everyone to have from my presentation here is three things I want you to learn from this, and one thing to forget.

So to start, the focus of this is on geospatial products at Google. Somebody might wonder, what is Google's relationship to geospatial products? And our core mission statement is to organize the world's information and make it universally accessible and useful. And implicitly, there's geospatial right there. It's the world's information, that's where all the data comes from, to some degree or another. And typically, by making it useful, that means easy to work with and scalable solutions because, as I'm sure most of the folks on this call know, geospatial data can be very large and difficult to work with.

So in practice, what does that mean? Well, this is the main thing that I hope you can take away from my talk today, that within Google Cloud, we have two tier types of capabilities. One is visualization tools, and the other is analysis tools. So by visualization tools, I mean tools that are good for a human to dynamically and iteratively explore geospatial data. Whereas, an analysis tool is more focused on, I'm doing science or research, large scale computation across geospatial data, not necessarily a consumer product. And there's a lot of overlap between these two. This isn't a very stark delineation by any means, but I think it's a useful framing for discussion.

So the first thing I want to start with, this is the thing to forget; Google Earth Enterprise, it's still around, we open sourced it in 2017. I talk to customers about it all the time. It has one of the most loyal fan bases of any consumer product ever. I've regularly met with folks who may or may not have been jailbreaking government devices so that they can continue to install it on their platforms, but that is no longer an actively, directly maintained product from Google, despite the fact that it's really influenced a lot of the other products that we have.

Maps is the other largely visualization capability. I think probably most people are familiar with maps from a consumer standpoint. There's a Maps app that a lot of folks have on their phone, but there's, that's actually a collection of APIs, and I'll dive into that in more detail next.

On the other side, we have the analysis tools. And the two of those there that I want to focus on today are Google Earth Engine, which is for raster based analysis. It has a long history of being supported in the scientific community and adoption there, in the nonprofits, and we've recently seen an uptick in adoption within government customers. And BigQuery GIS, which is more for vector-based analysis, and BigQuery is our data warehouse. But to start, I'm going to go ahead and jump into Maps and the associated APIs there.

So within maps, there's three core areas of APIs. There's the Map's map's API, the routes APIs, and the Places APIs. And I think you'll find that most of these are fairly straightforward. Maps are typically about different ways of visualizing geographic data. Routes are APIs to help to map traffic over that, and do directions, and transportation, and Places are to tell you things about an individual place.

So the Map's APIs receive 25 million updates a day. That's over 289 updates a second. And these are the traditional Maps that you might've experienced on web apps or on different webpages when you're looking up local restaurants and things like that. There's different ways that those can be embedded into different technologies. And those also include our street view APIs, so if you want to actually see a street level view of a location, it enables that.

The Routes APIs are for things like directions or fleet routing. It could be an point to point directions for an individual, or if you have a large number of vehicles and a large number of stops that they need to go to, you could think like a fleet of delivery trucks, for instance, we have a distance matrix API that assists with that.

And then the Roads API gives you metadata about any individual road. So if you're on a, if you wanted to create an application that was saying like, hey, what is the speed limit for this given road? Or something like that, you could source that information from that API.

And then the third type of APIs that we have under Maps, so the Places APIs, we have data on over 150 million different places around the globe, and we add about 700,000 places a month to that data set. The Places API gives you information about a building, oftentimes it includes a street view image if you want to source that, as well. If it's a place of business, it can give you operating hours or busy-ness information. And then some of another one of these APIs that receives a lot of attention is the Geocoding API. So converting from an address to geographic coordinates or vice versa.

One other recent, relatively recently announced capability is our 3D Tiles, API. So while the old Google Earth Enterprise application is no longer supported, you'll find that a lot of the things that folks enjoy doing with it, being able to zoom around cities and see them in 3D, the 3D Tiles API supports a lot of those capabilities. I'm very excited about that. We've been partnered with, it's in pre GA right now. It's not quite general availability, but if you're interested in this, please follow up with me afterwards and we can get you involved. We've also been working with CSM on it a little bit to have a framework for displaying that.

All right. So I briefly covered the visualization tools at a high level, and so now I'm going to dive into the analysis tools. So that's Earth Engine and BigQuery GIS. So quickly, I mentioned this in passing, this might be old news for everyone on the call who's pretty invested in GIS and those capabilities, but in case you're not, commonly, geospatial data comes in two different forms. That would be vector data, data and raster data. Typically, vector data is tabular data, it's machine readable descriptions of geometric features. Typically, those are points, lines, polygons, things like that. Some examples that we just saw from census include roads, or census tracks, or building footprints, or the boundaries of accounting. Those are all things that would typically be represented as vector data.

On the other hand, we have raster data. The archetypal example of that would be imagery, anything that you would represent as pixels, but it doesn't have to be just imagery. It could be soil moisture, it could be weather, it could be weather data sets, such as surface temperature of water, anything that you could conceivably represent as pixels. And typically, a lot of these data types, even for imagery, they have the components of that data separated into what are called different bands. So a simple example might be an image is composed of RGB, red, green, and blue, and with Earth Engine, you can actually analyze each of those bands individually, not just for imagery, but any dataset that it's provided as bands.

Okay, but first I'm going to talk about Earth Engine, which, like I just said, is for our raster, is for analyzing raster data at scale. So what is Earth Engine? That is a cloud-based platform from Google that allows you to analyze global data sets of raster data, typically satellite imagery, in the cloud. So as I'm sure everyone on this call is aware, the earth is not a static place. It's constantly undergoing change, both naturally and manmade. And via satellite imagery and other raster data sources, you can look at how those that are taken over time, you can see how the surface of the earth is changing over time. And so Google Earth Engine gives you a tool to look at those large geospatial raster data sets and look at them over time for a given area of interest and analyze them. Analyze the different bands of those images or other data sets. We have over 80 petabytes of geospatial data already built into the platform. Hold on one second. My children are using the Alexa announcement feature at an inopportune time. So it has over 80 petabytes of geospatial data already built into it, and it's been used by over a hundred thousand scientists around the world.

So just to try to anchor this a little bit more, what are the kinds of things you can do with those geospatial data? So a colleague of mine had the question, how much grass grows in the United States? And he was asking this question in 2021, and using Earth Engine, he was able to solve that by taking satellite imagery and using Earth Engine to analyze it at scale, and he built an entire app called, you can go to it right now, well, after my talk, and go to rangelands.app and check it out. And this was when he was employed by the government, and he was able to do that for every 16 days from 1986 to the present. And so he learned that in 2021, there were 596 million tons of grass that grew in the United States. And the neat thing about it is, once you can do that analysis, you can do it retrospectively going back 20, 30 years and see, oh, why is it varying over time? Oh, these years where it's low is because of drought, et cetera, et cetera.

So the way Earthen Engine works, and I kind of touched on this a little bit, is that partly, there's a data catalog component to it and a computational platform component to it. So first, I'm going to dive into the data catalog side. So within the data catalog, there's public data sets, that's that 80 petabytes that I talked about before, as well as private data. So you can bring in your own geospatial raster data, typically geo tif, but we support other types, and it remains yours, it doesn't get shared with anyone unless you specifically want to share it with another group on your team or a party you're collaborating with. But you're able to leverage the public data that we've ingested in your analysis.

And so within the public data, you can look that up within our Earth engine data catalog. And it goes, we have over 50 years of data going back to, I think, the early Landsat imagery. And there's different types. There's all sorts of different types of data in our archives, climate and weather imagery, geophysical, and you can check out that full catalog on the web. If you go to the website, it looks something like this, and you can explore that and look into it.

So for example, one of the data sets that we have is from MODIS from NASA. They have two satellites that are imaging the earth at a 500 meter resolution. And one of the products from that is a last snowfall data set, and using Earth Engine, you can very simply explore that. So if you look that up in our data catalog, you're going to see a website that looks like this that provides some metadata about the data set, the name, the availability. So we see that this one goes back to 2000 and it's been updated up until every day when I made these slides, the source and a brief description of it, there's a description of the bands, and it's not shown in this image, but there's a little code snippet that you can use to get started with it right away. And if you load that code snippet and click on it and open it up in Earth Engine, it looks something like this. You can go ahead and run it and start exploring it right away.

I actually modified this slightly. I changed the default dates for the example to early February, 2010. For folks in the DC area who were living here at that time, that was one of our fun snow events. And so I'm looking here, at the snow cover from one of those snowpocalypses. And this is an interactive coding window that you can use. This is where you can view the results of that, and over here we have things like the different scripts and data assets that I have access to.

There are a large number of different data sets within the catalog. If I had more time, I'd dive into these in more detail, but some of the ones that I find pretty interesting are from NOAA, it's the Global Forecast System data. We actually have large scale international boundary polygon data. That's very useful for if you want to constrain a particular analysis to a geographic region or a country, it's very easy to bring that in, so you'll see that in a lot of the demos.

And then another one that I think is really cool is one that's actually provided by Google. Most of the data sources that we have, Google is ingesting those and maintaining those from public domain sources, typically the government, but we do generate and share a few of our own, and this is one of them. That's the Google Dynamic World Land Use data set, which is really cool and fun to look at. And we're calculating, I think, eight different land use categories across the globe.

So coming back, like I said before, part of Earth Engine is the 80 petabytes of data that are analysis ready for you to start using. And the other side, the other component of that is a computational platform. So when we say a computational platform, let me explain what that means a little bit. So there's two aspects to that. There's on the fly computation and batch computation. The on the fly computation is what I was just showing you via that browser-based interface, and batch computation is more for when you... Typically, the way a user works with it is they iterate in the JavaScript browser, I'm doing the on the fly computation, figure out the analysis that they want, tweak it, and then perform a batch computation analysis to run that at a much larger scale.

And one common thing I always like to clarify is that there's a client server relationship here. Typically, what you're doing is on your laptop or on your local desktop, you're writing the code, but all of these large scale analysis, the large scale computation, that's taking place in the Cloud, you're able to leverage the power of Google's cloud and our parallelism that we can achieve there to make it much more responsive. So what you're actually doing is you're submitting the request, the description of the analysis you want to perform to the Cloud, it's being performed there, and then you're retrieving the results back.

I already showed you this. This is what the interactive code editor looks like. It's browser based, you can write your code in the middle, so you'll see the results of that after you run it in the visualization window. And there's all sorts of baked in, all of your scripts are managed there, it's a GIT based system, so you can do version controls, you have sharing capabilities and any sort of custom assets that you've uploaded would show up there.

In addition to the JavaScript programming SDK, we also have a Python SDK. So if you're a fan of Jupiter Notebooks, it integrates very well into that environment. And then there's also a rest based capability, although that's not used quite as often.

Just to dive into a little bit more detail about what's happening under the hood on the server or the Cloud side, when we ingest data from our data pipelines where you upload your own data, those images are tiled in down sample to create an image pyramid. And then during computation, what we're able to do is actually constrain the computation to the appropriate geography and resolution, the downsampling level for the particular analysis, and do it in parallel because it's been tiled, so that the end result there is that we're able to perform analytics much faster using a cloud-based system.

By analytics, there's a lot of different things that you can do with Earth Engine. I don't have time to really dive into it in great detail. The classic example that if you, like Hello, World! is calculating NDVI, if that makes sense. But at a high level, you can sub-select bands or resolutions or projections of an image, you can apply an algorithm to that image. There's a library of functions that we have readily available. You can write your own. You can filter a collection of images by time and space or metadata. You can then apply your algorithm or algorithms to all of those filtered images, then combine them down into a single aggregate image and then perform computational statistics, aggregate statistics across that derived image.

So once someone has gone through the exercise of performing a useful calculation, they can, then, incorporate that into what we call an app, and then very easily turn that into a web facing tool that others can use. So one example is one from Ukraine, crop production, looking at how wheat is being grown across Ukraine going back several years. Another one from the UN, here, is flood mapping. And these are, once you have your analysis, you can very easily turn that into an app that others can utilize. And then it also integrates with the rest of Google Cloud if you want to do more advanced AI and ML within Earth Engine.

So the last thing I'll touch on very quickly is BigQuery GIS, which, like I said before, is our vector-based analytics solution. A classic example that you might want to use with that is, hey, I have over 12 million statewide parcels across Texas, how many of them have impervious surfaces? Using BigQuery GIS, my colleague was able to calculate that result in 12.1 seconds, whereas, performing the similar analysis using Cloud SQL took over five hours.

I didn't really touch on this, but Google Cloud, we're a Cloud service provider. We have data centers around the United States with a whole suite of Cloud service capabilities that you would expect. BigQuery is one of our core big data capabilities that acts as a data warehouse. It has a similar architecture model to Earth Engine, in that storage and compute are handled separately. But really, it's a linchpin for a lot of the analytics that you might want to perform, both in terms of being able to ingest data from different sources and analyze it with a variety of tools. We have a number of public data sets from a lot of government customers, including Census, and we have native support for GIS data, GIS vector data and geography types, via the BigQuery GIS capabilities, and that includes native support for your ST_functions. And there's a number of different visualization tools that you can utilize with it, both your own third party ones, as well as something called BigQuery Geo Vis, or it integrates well into our Data Studio and Looker capabilities.

Like I said before, there's not, necessarily, clear delineations between all of these different capabilities. There is some overlap, but generally speaking, if you're starting from a raster based analytic, then you would start with Google Earth Engine, whereas, if you're dealing predominantly with vector data, you should explore BigQuery first.

And so thank you very much for your time today, and I hope you take away the one thing... You leave behind the one thing I hope you forget for right now, and take with you the three other capabilities that we have. If you'd like to reach out, if you have any other follow-up questions or just want to chat, here's my contact info. Thank you.

Carten Cordell:

Thanks, Aaron. Now let's hear from Dorothy Spears-Dean, Deputy State Coordinator of Virginia's Department of Emergency Management on the critical and expanding role of GIS and next generation 911. Dorothy.