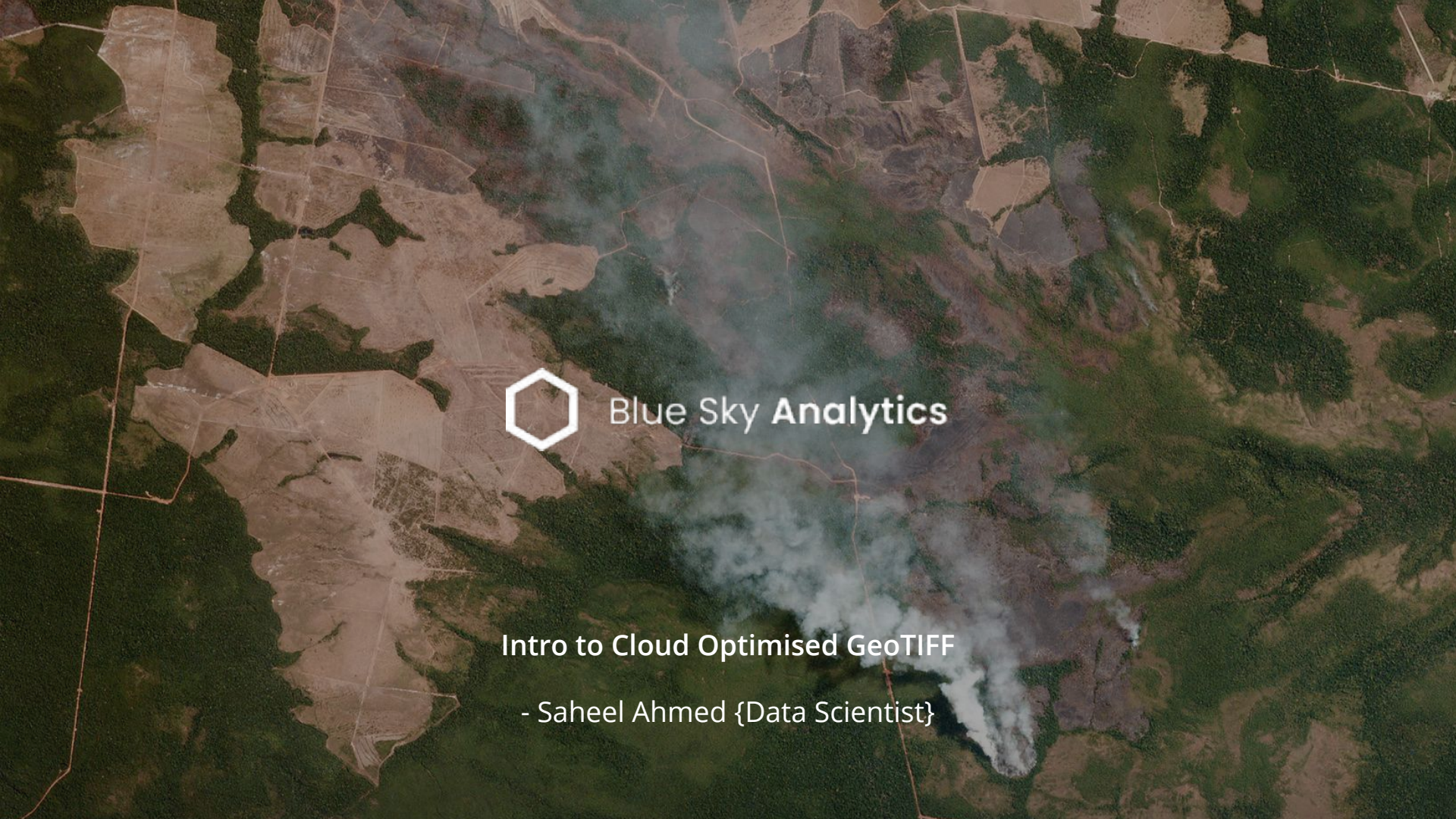
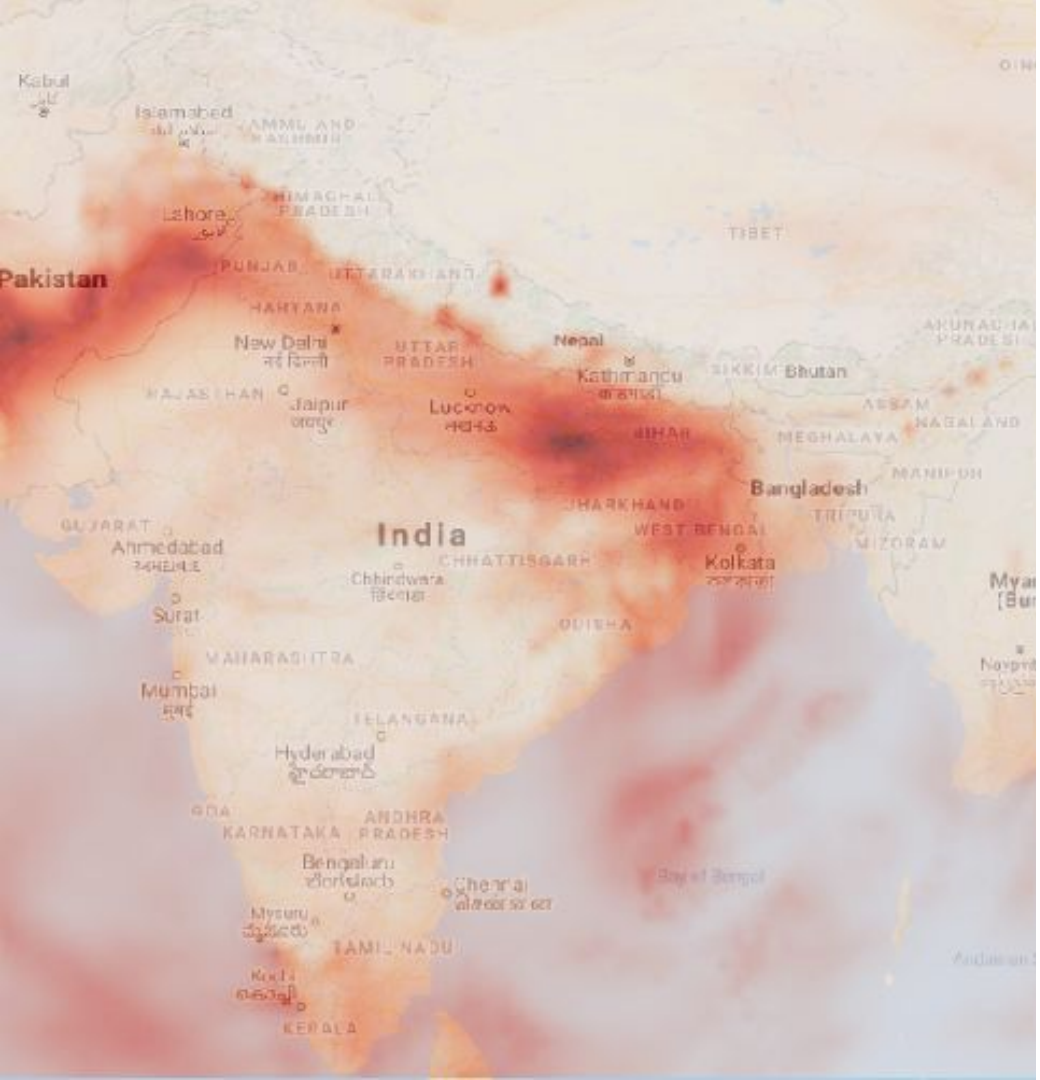


Intro to Cloud Optimised GeoTIFF

- Saheel Ahmed {Data Scientist}



Blue Sky **Analytics**

BlueSky is building cutting-edge technology that can identify environmental parameters like air pollution, fires, size of water bodies etc. using satellite imagery.

We use AI to analyze this data and examine trends over space & time.

By combining the power of AI & Satellites, we hope to build the equivalent of a ‘Bloomberg for Environmental Data’

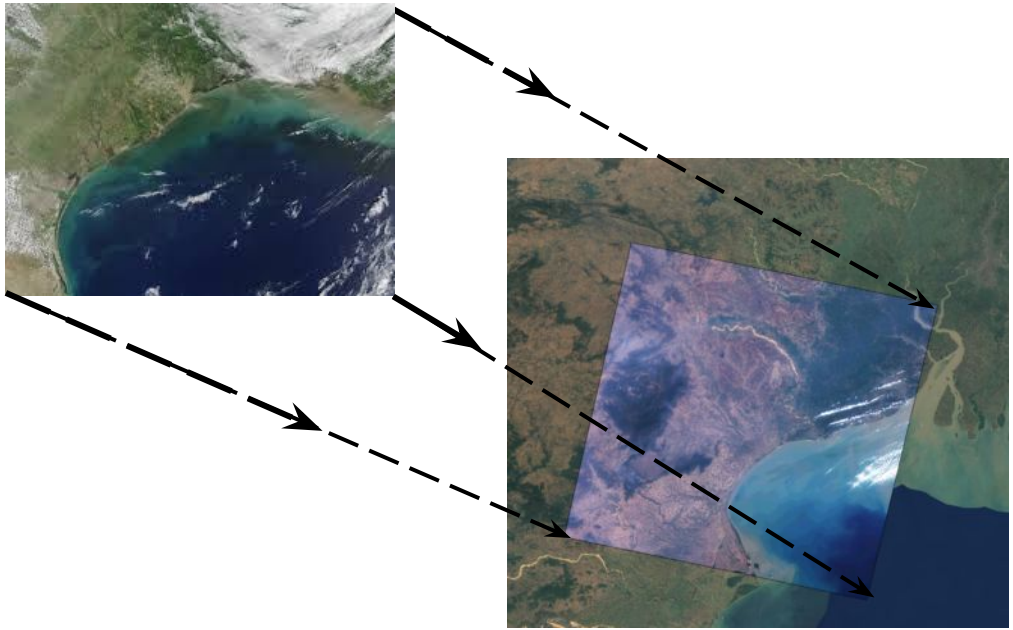
MIT SOLVE

esa

TIFF & GeoTIFF

Tagged Image File Format (TIFF)

- Bitmap raster file
- Lossless data compression
- Edit & Resave with no compression loss

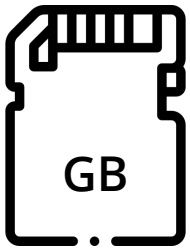


GeoTIFF

- TIFF with embed georeferencing
- Tells you where on earth this image is from.

Blue Sky Analytics

Pain Points in Handling Geo-spatial Data

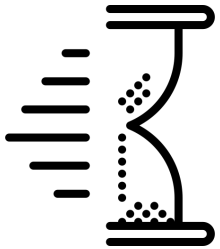


HEAVY FILE SIZE OF SATELLITE DATA

Each image strip can be 20-40 GBs. Moving it around is a pain.

MINIMIZE DATA DUPLICATION

Copy and cache is a pain. Need single source of truth for many softwares.



REAL TIME & DYNAMIC PROCESSING

Image processing on the fly based on user needs & requests is a pain.

Cloud Optimized GeoTIFF (COG)

GeoTIFF + INTERNAL ORGANIZATION

A GeoTIFF file, aimed at being hosted on a HTTP file server, with an internal organization like metadata → full res image → overview metadata → final imagery. Compression + Internal organization like tiling & overviews are best practices for software to quickly access imagery.

&

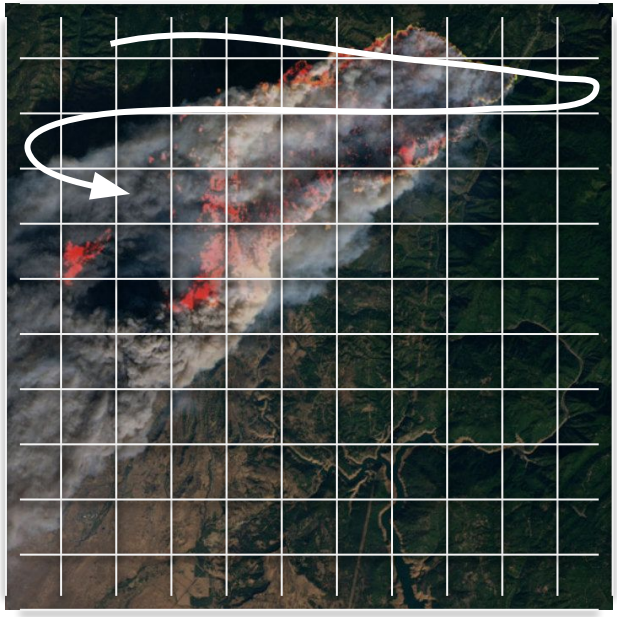
HTTP GET RANGE

Optimized cloud workflows by allowing clients issuing HTTP GET range requests to ask for just the parts of a file they need. Go from smallest overview → full resolution for streaming & progressive rendering.

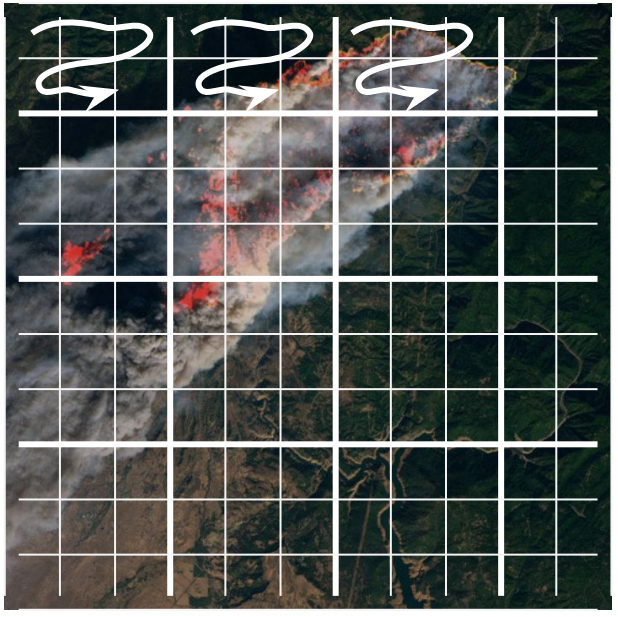
Internal Tiling

With tiles much quicker access to a certain area is possible, so that just the portion of the file that needs to be read is accessed.

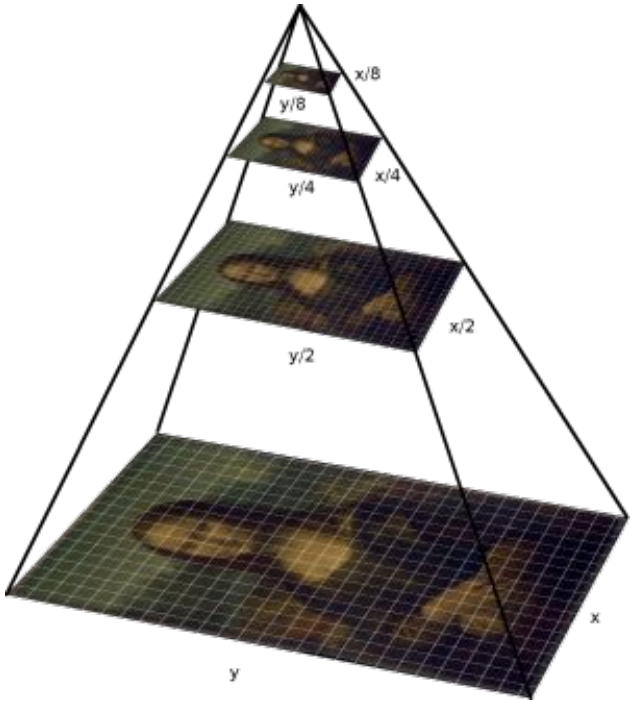
REGULAR GeoTIFF: STRIPE {2000 X 1}



COG: TILING {256 X 256}



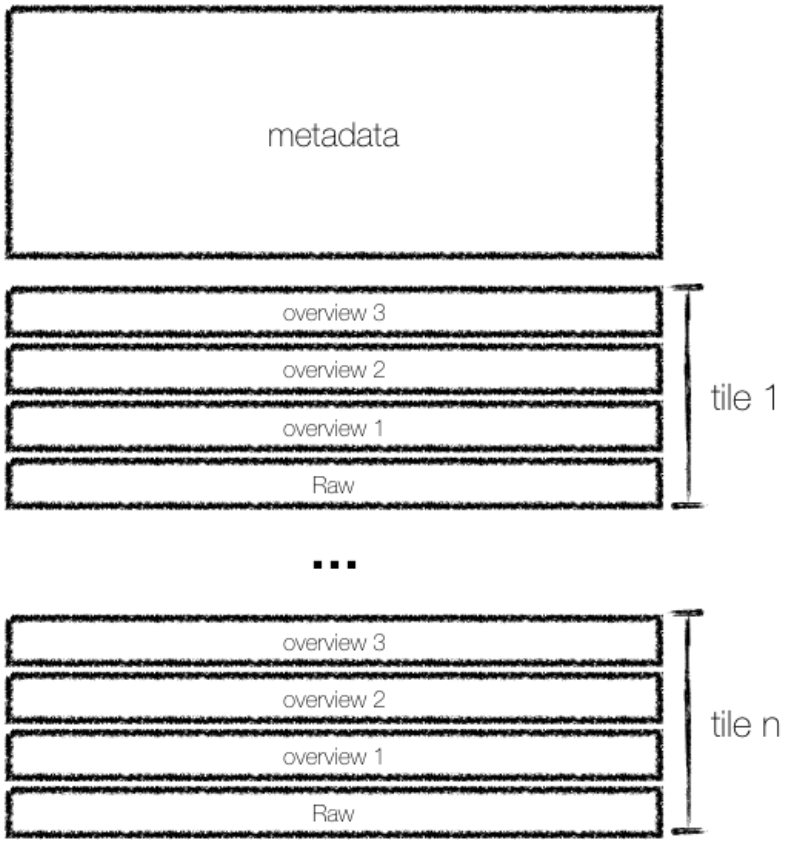
Internal Overview



- Overviews create downsampled versions of the same image.
- ‘Zoomed out’ versions of the original image. Lesser detail & smaller size.
- GeoTIFFs usually have multiple overviews, to match different zoom levels.
- Adds to overall file size, but is served much faster.

Cloud Optimized GeoTIFF (COG) Structure

Compression + Internal organization like tiling & overviews are best practices to enable software to quickly access imagery.

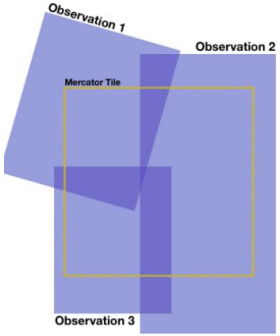


Blue Sky Analytics

HTTP GET range requests

- Introduced in HTTP Version 1.1. Comes into play in GET requests, when a client is asking a server for data.
- Byte Serving: If the server advertises `Accept-Ranges: bytes` header in its response, client can request just the bytes that it needs from the server.
- Very useful for serving things like video, to allow playing without download the entire file.
- Range requests are an optional field.
- Almost all the object storage options on the cloud (Amazon, Google, Microsoft, OpenStack etc) support the field on data stored on their servers.
- Almost any data stored on the cloud is automatically able to serve up parts of itself, as long as clients know what to ask for.

Power of Cloud Optimized Geo-TIFFs



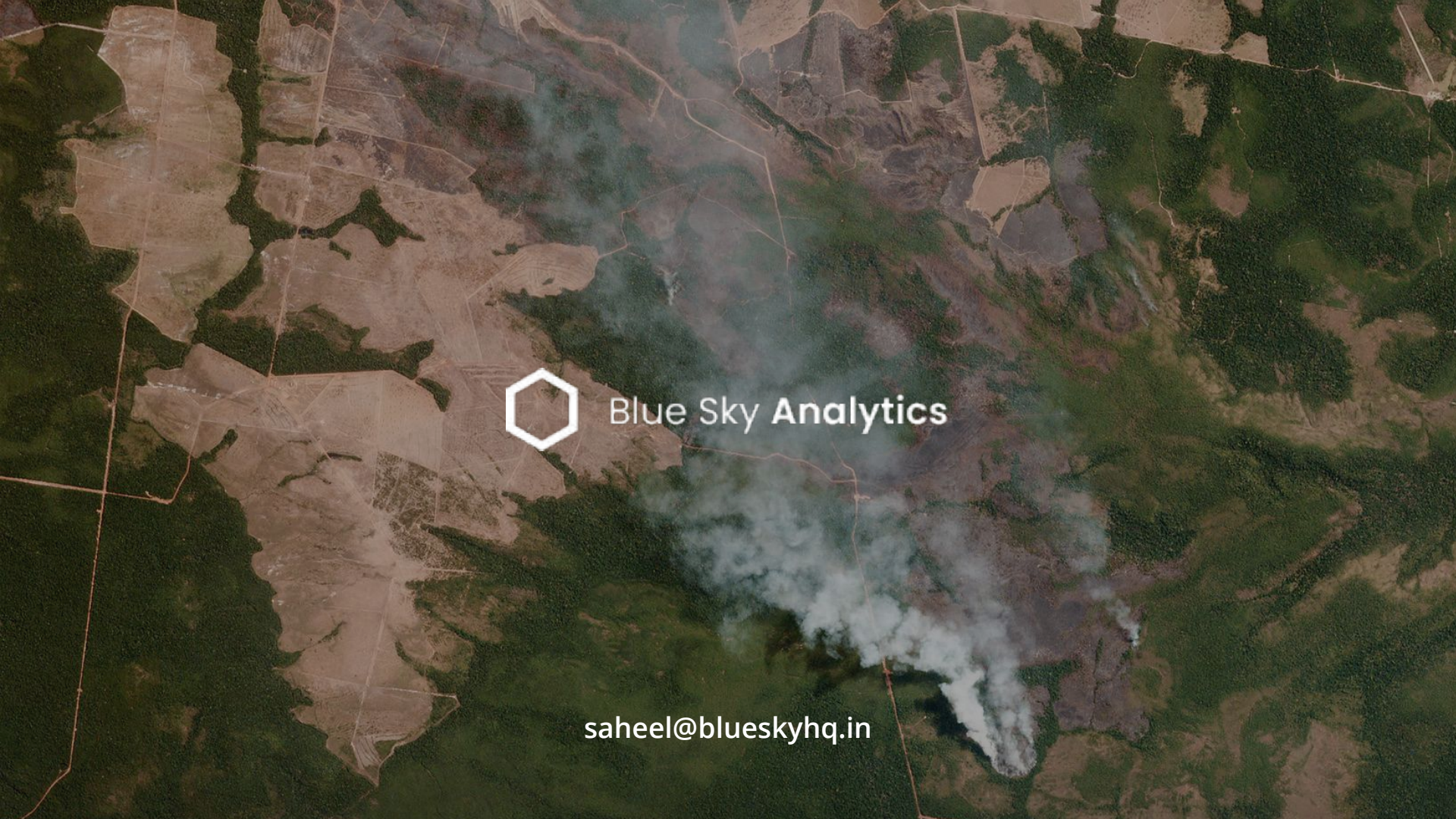
OVERLAY MOSAICs ON THE FLY

GENERATE & LOAD DYNAMIC TILES

Converting Regular Geotiff into a Cloud Optimised Geotiff.

- Given an input dataset input.tiff with already generated internal or external overviews, a cloud optimized GeoTIFF can be generated with:
    - gdal_translate input.tiff output.tiff -co TILED=YES -co COPY_SRC_OVERVIEWS=YES -co COMPRESS=LZW

- If the files doesn't have overviews run the below command and then the above command:
    - gdaladdo -r average input.tiff 2 4 8 16



saheel@blueskyhq.in