Measuring Vegetation Health to Predict Natural Hazards

Suneel Marthi

June 18, 2019
Berlin Buzzwords, Berlin, Germany

$WhoAmI

Suneel Marthi

@suneelmarthi

- Member of Apache Software Foundation
- Committer and PMC on Apache Mahout, Apache OpenNLP, Apache Streams

Agenda

- Introduction
- Satellite Image Data Description
- Satellites and Vegetation
- What is NDVI ?
- Feature Engineering
- Cloud Classification
- Vegetation Prediction
- Apache Beam
- Beam Inference Pipeline
- Future Work

Introduction

Deep Learning has moved from Academia to Industry

Availability of Massive Cloud Computing Power

Combination of Compute Resources + Big Data with Deep Learning models often produces useful and interesting applications

Introduction

Computer Vision for Satellite Imagery

Availability of low cost satellite images for research

Train a Deep Learning model to Determine Vegetation Health

Goal: Identify Vegetation Index from Landsat8 Images

Data: Landsat8

Earth observation mission from EOS

10 spectral bands, from RGB to SWIR (Short Wave Infrared)

Spatial resolution: 30m/px (Visible, NIR bands)

Free and open data, updated each day

Landsat8 Payload

Operational Land Imager (OLI)

Thermal Infrared Sensor (TIRS)

Resolution of 30m/pixel(visible, NIR, SWIR)

Landsat8 Bands

Band	Range(nm)	Resolution	Spectrum
B1	435-451	30m	Coastal
B2	452-512	30m	Blue
B3	533-590	30m	Green
B4	5636-673	30m	Red
B5	851-879	30m	NIR
B6	1566-1651	30m	SWIR-1
B7	2107-2294	30m	SWIR-2

How is this related to Vegetation?

Photosynthetically Active Radiation (PAR) 400 - 700nm range

Plants absorb solar radiation in PAR region

Plants re-emit solar radiation in NIR region

Hence, plants appear bright in the NIR spectrum

https://en.wikipedia.org/wiki/Near-infrared_spectroscopy#Agriculture

How is this related to Vegetation? (Contd)

Plant Chlorophyll absorbs visible light (0.4 to 0.7μm)

Plant leaves reflect NIR (0.7 to 1.1μm)

More plant leaves, greater the radiance emited from plants

https://en.wikipedia.org/wiki/Near-infrared_spectroscopy#Agriculture

Difference in plant reflectances can help determine their distribution in satellite imagery

https://en.wikipedia.org/wiki/Near-infrared_spectroscopy#Agriculture

How do we measure this phenomenon?

Normalized Difference Vegetation Index (NDVI)

Used to monitor agriculture production, drought, hazardous fire zones

$$NDVI = \frac{(NIR - Red)}{(NIR + Red)}$$

Bands 5 and 4 in Landsat8 images

NDVI Range: [-1,1]

Dense Green Vegetation canopy is typically 0.3 - 0.8

Standing water like Deep Oceans (away from coast) typically have -ve or small +ve values

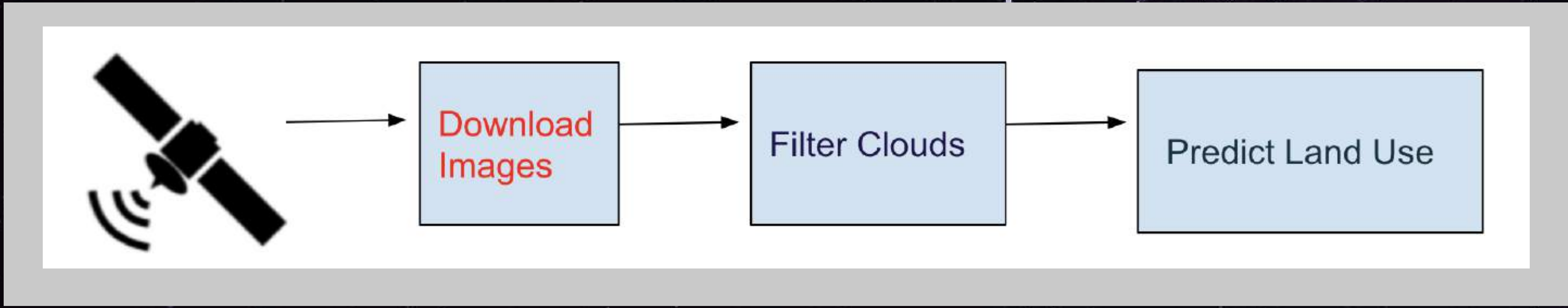
Soil with a high NIR, tends to have a small positive value < 0.2

https://en.wikipedia.org/wiki/Near-infrared_spectroscopy#Agriculture

Cover	Red	NIR	NDVI
Dense Vegetation	0.1	0.5	0.7
Dry Soil	0.269	0.283	0.025
Clouds	0.227	0.228	0.002
Snow and Ice	0.375	0.342	-0.046
Water	0.022	0.013	-0.257

https://grindgis.com/blog/vegetation-indices-arcgis#1

Workflow



Goal: Identify Vegetation Index from Landsat8 Images

Data Acquisition

Images downloaded from AWS Earth (available as Open Data)

Data refreshed continually

```
aws s3 sync s3://landsat-pds/c1/L8/044/033 ~/Downloads/landsat-04
```

Data

Registry of Open Data on AWS

aws

Landsat 8

earth observation | satellite imagery | geospatial | natural resource | sustainability | disaster response

Description

An ongoing collection of satellite imagery of all land on Earth produced by the Landsat 8 satellite.

Update Frequency

New Landsat 8 scenes are added regularly as soon as they are available.

License

There are no restrictions on the use of data received from the U.S. Geological Survey's Earth Resources Observation and Science (EROS) Center or NASA's Land Processes Distributed Active Archive Center (LP DAAC), unless expressly identified prior to or at the time of receipt. More information on licensing and Landsat data citation is available from USGS.

Documentation

https://docs.opendata.aws/landsat-pds/readme.html

Managed By

planet.

See all datasets managed by Planet.

Contact

https://lists.osgeo.org/mailman/listinfo/landsat-pds

Usage Examples

- A Gentle Introduction to GDAL Part 4: Working with Satellite Data by Planet
- Aggi: the fire detecting Twitter bot by Aggi
- Apps for exploring and analyzing Landsat imagery on the fly by Esri
- COG-Explorer - View Cloud Optimized GeoTIFF images in the browser directly from object storage by EOX
- Development Seed Geolambda by Matthew Hanson
- EOS Land Viewer by Earth Observing System

Resources on AWS

Description
Scenes and metadata

Resource type
S3 Bucket

Amazon Resource Name (ARN)
`arn:aws:s3:::landsat-pds`

AWS Region
`us-west-2`

Description
New scene notifications

Resource type
SNS Topic

Amazon Resource Name (ARN)
`arn:aws:sns:us-west-2:274514004127:NewSceneHTML`

AWS Region
`us-west-2`

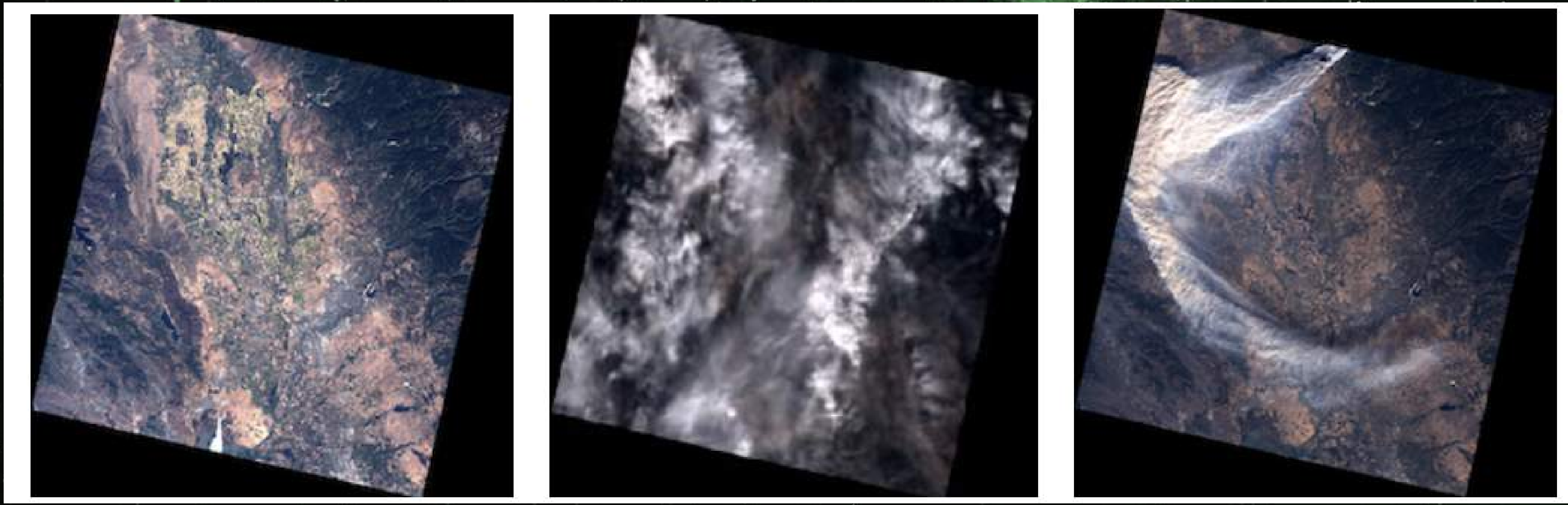
Description
S3 Inventory (ORC)

Resource type
S3 Bucket

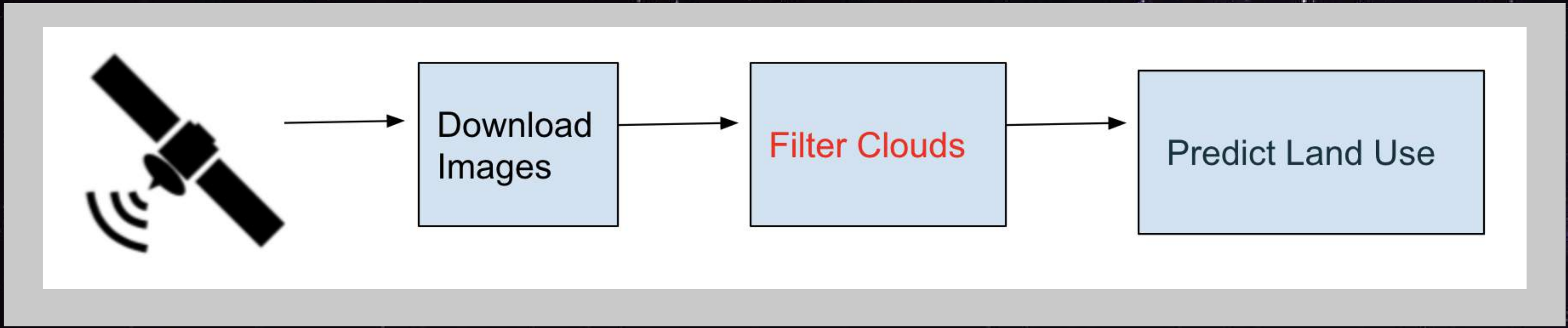
Amazon Resource Name (ARN)
`arn:aws:s3:::landsat-pds-inventory`

AWS Region
`us-west-2`

Data



Workflow



Filter Clouds

Need to remove cloud cover from images before further processing

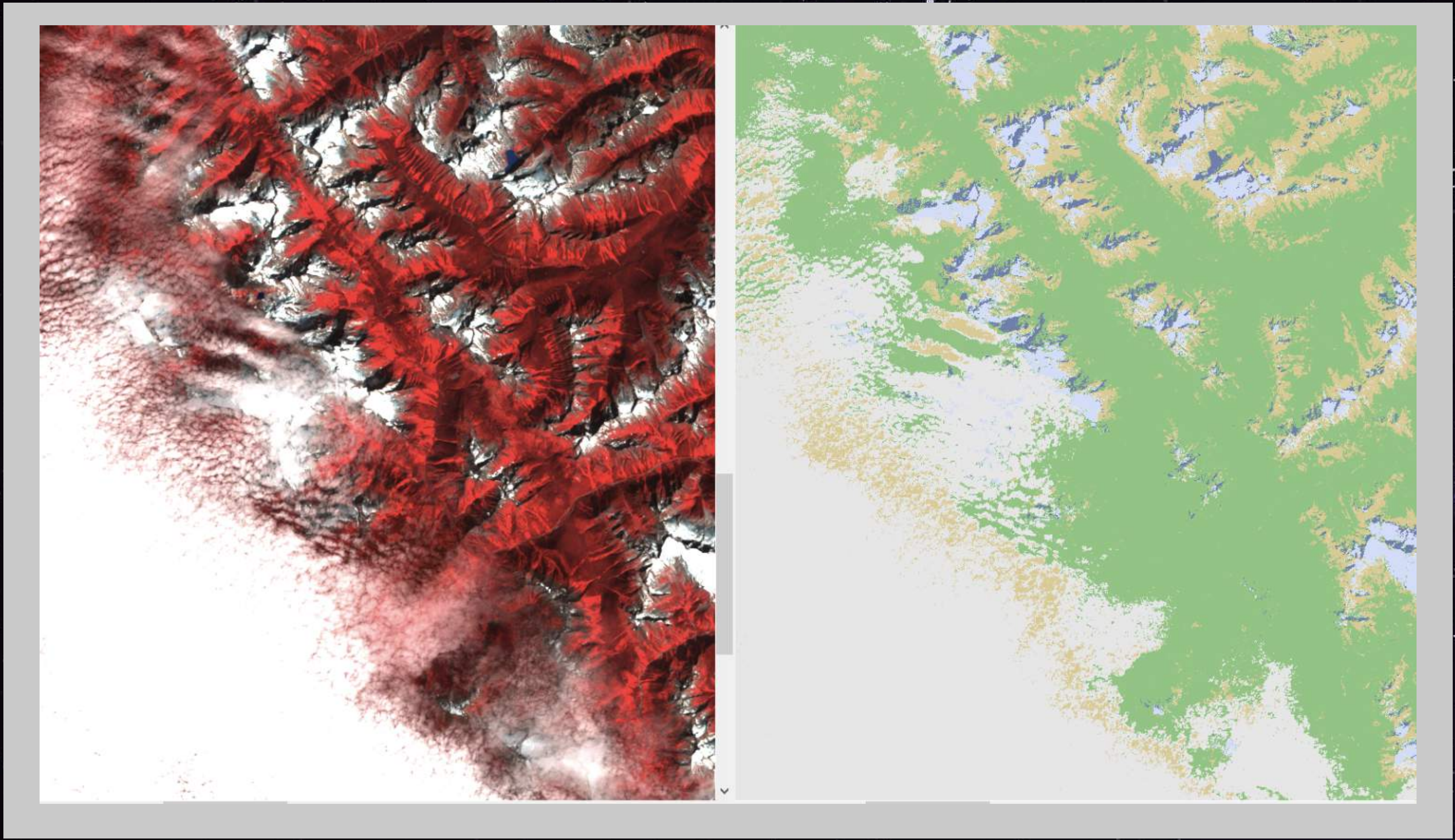
Approach: Mask the clouds from all the feature bands

Compute Median Cloud Masking Composite for the year

Cloud Masking Function

```
/**
 * Function to mask clouds based on the pixel band of Landsat 8 S
 * @param {Image} image input Landsat 8 SR image
 * @return {Image} cloudmasked Landsat 8 image
 */
function maskL8sr(image) {
  // Bits 3 and 5 are cloud shadow and cloud, respectively.
  var cloudShadowBitMask = (1 << 3);
  var cloudsBitMask = (1 << 5);
  // Get the pixel QA band.
  var qa = image.select('pixel_qa');
  // Both flags should be set to zero, indicating clear condition
  var mask = qa.bitwiseAnd(cloudShadowBitMask).eq(0)
               .and(qa.bitwiseAnd(cloudsBitMask).eq(0));
  return image.updateMask(mask);
}
```

Example Results (B2, B3, B4 bands)



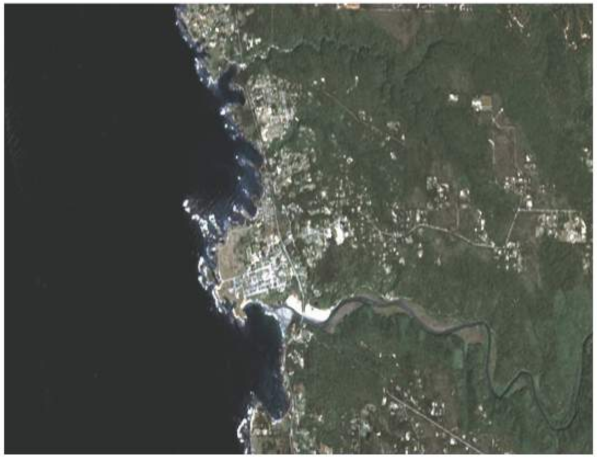
Feature Engineering

- Use Bands B2, B3, B4, B5, B6, B7 for training and prediction
- Apply the Cloud Masking function to each of the bands

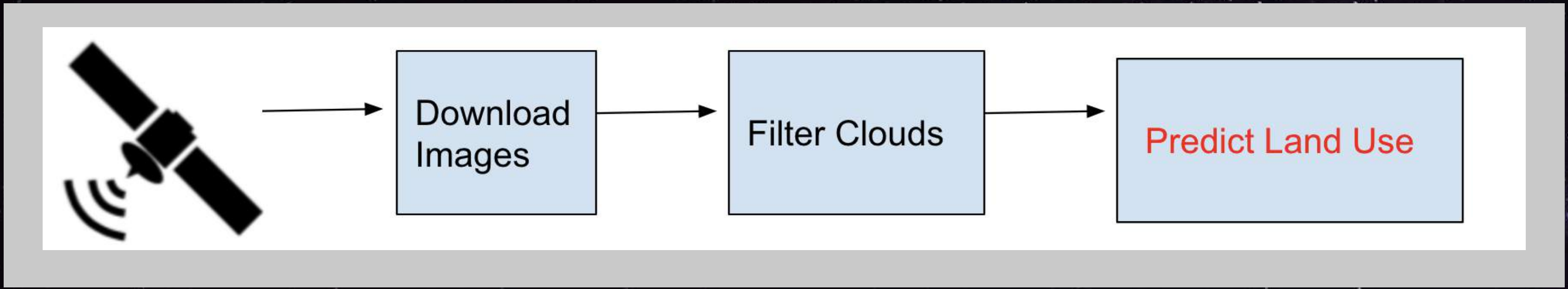
Feature Engineering (Contd)

- Use Landsat8 Surface Reflectance data for computing cloud mask
- Result: Each pixel is classified as 1 of {Vegetation, Barren, Snow/Ice, Water, Cloud}
- Ignore the Cloud pixels
- Compute NDVI from Bands B4 and B5, add to the feature vector

Prediction Labels



Workflow



Model Training

Two Dense Layers with ReLU and Softmax activations

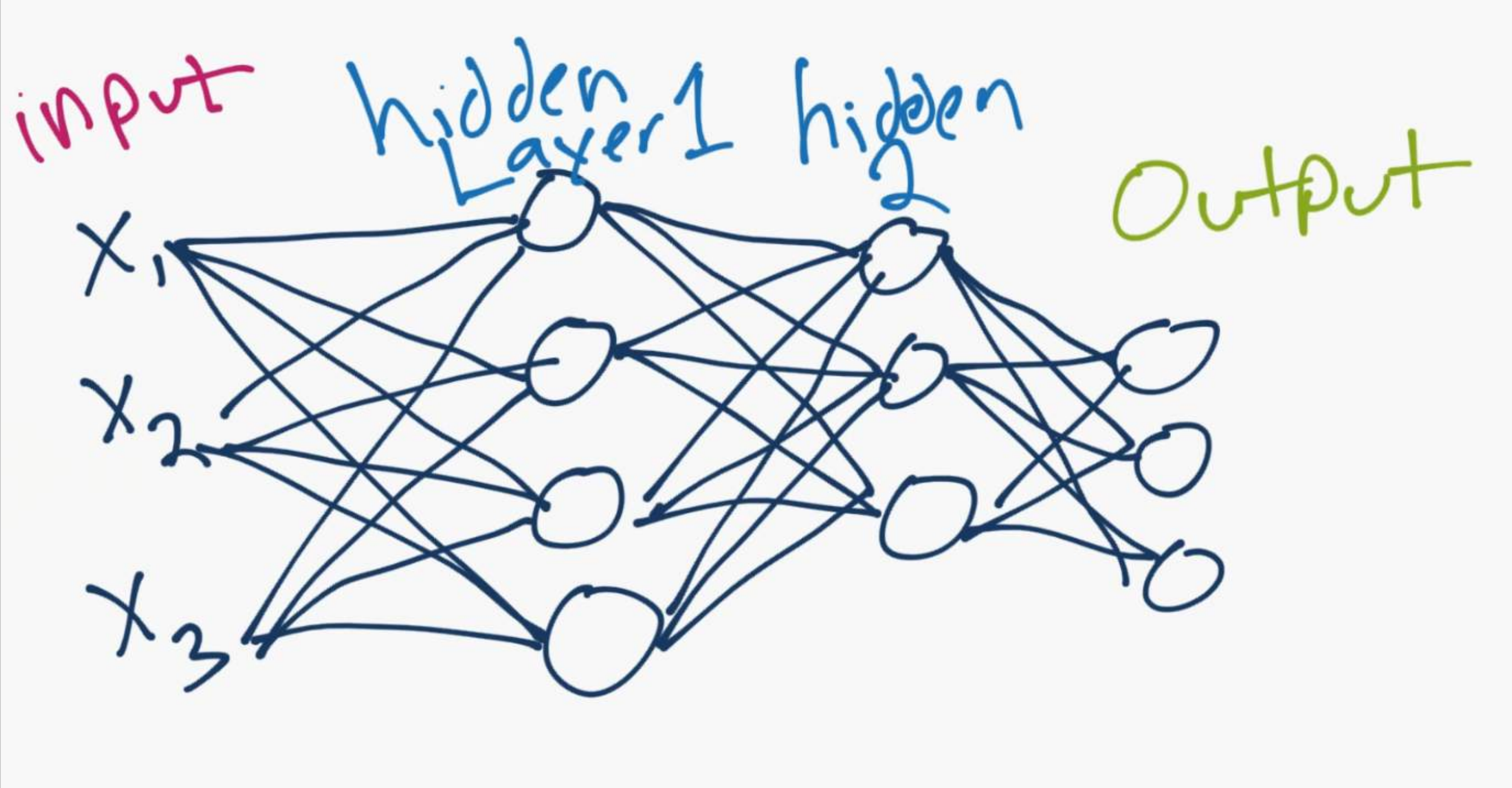
Dropout: 0.2

Optimization: Adam

Loss Function: Cross Entropy

One-Hot Encode the Labels

Model Training



Model Training - Keras

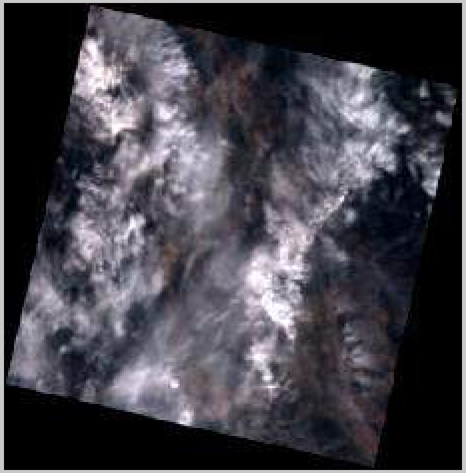
```
# Add the model layers
model = tf.keras.models.Sequential([
  tf.keras.layers.Dense(64, activation=tf.nn.relu),
  tf.keras.layers.Dropout(0.2),
  tf.keras.layers.Dense(nClasses, activation=tf.nn.softmax)
])

# Compile it with the optimizer and loss function
model.compile(optimizer=tf.train.AdamOptimizer(),
              loss='categorical_crossentropy',
              metrics=['accuracy'])
```

Paradise, CA - Landsat8, October 7, 2018

NDVI: -0.23979

Paradise, CA - Landsat8, October 23, 2018

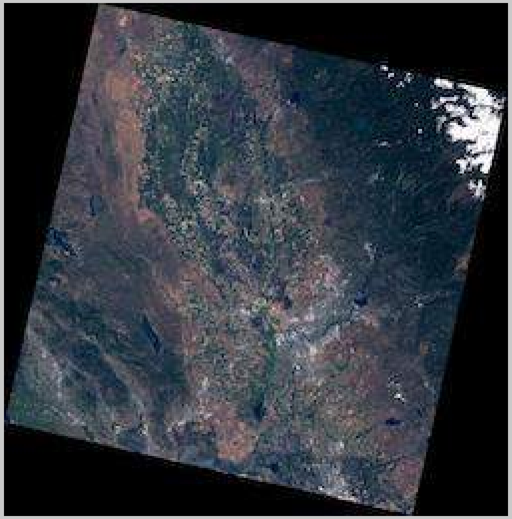


NDVI: -0.2407

Paradise, CA - Landsat8, November 8, 2018

NDVI: -0.23984151008195187

Paradise, CA - Landsat8, June 4, 2019



NDVI: -0.24

Scaling it all Out

How to Scale – Batch or Stream ?

"Batch is an extension of Streaming, except when Streaming is an extension of Batch"

-- Shannon Quinn, Apache Mahout

Spark or Flink ?

"Spark Streaming is for people who want to operate on their streams using Batch idioms.
Flink Batch is for people who want to operate on their batches using Streaming idioms."

-- Joey Frazee, Apache NiFi

Was ist Apache Beam?

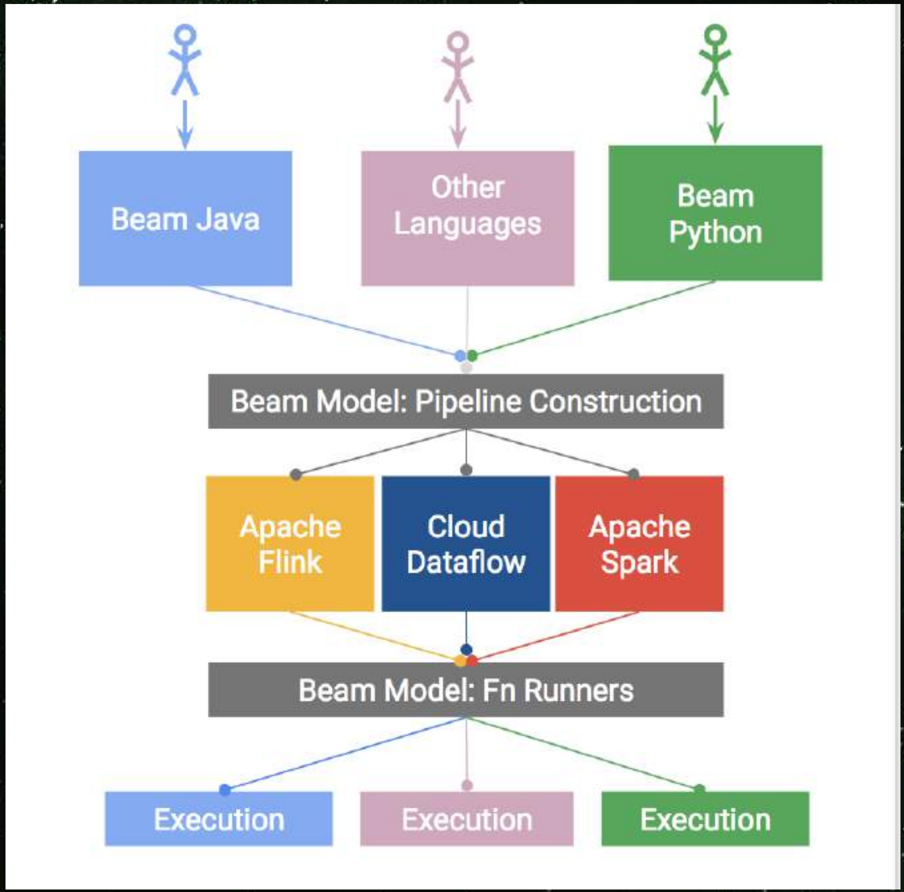
- Agnostic (unified Batch + Stream) programming model
- Java, Python, Go SDKs
- Runners for Dataflow
  - Apache Flink
  - Apache Spark
  - Google Cloud Dataflow
  - Local DataRunner

Warum Apache Beam?

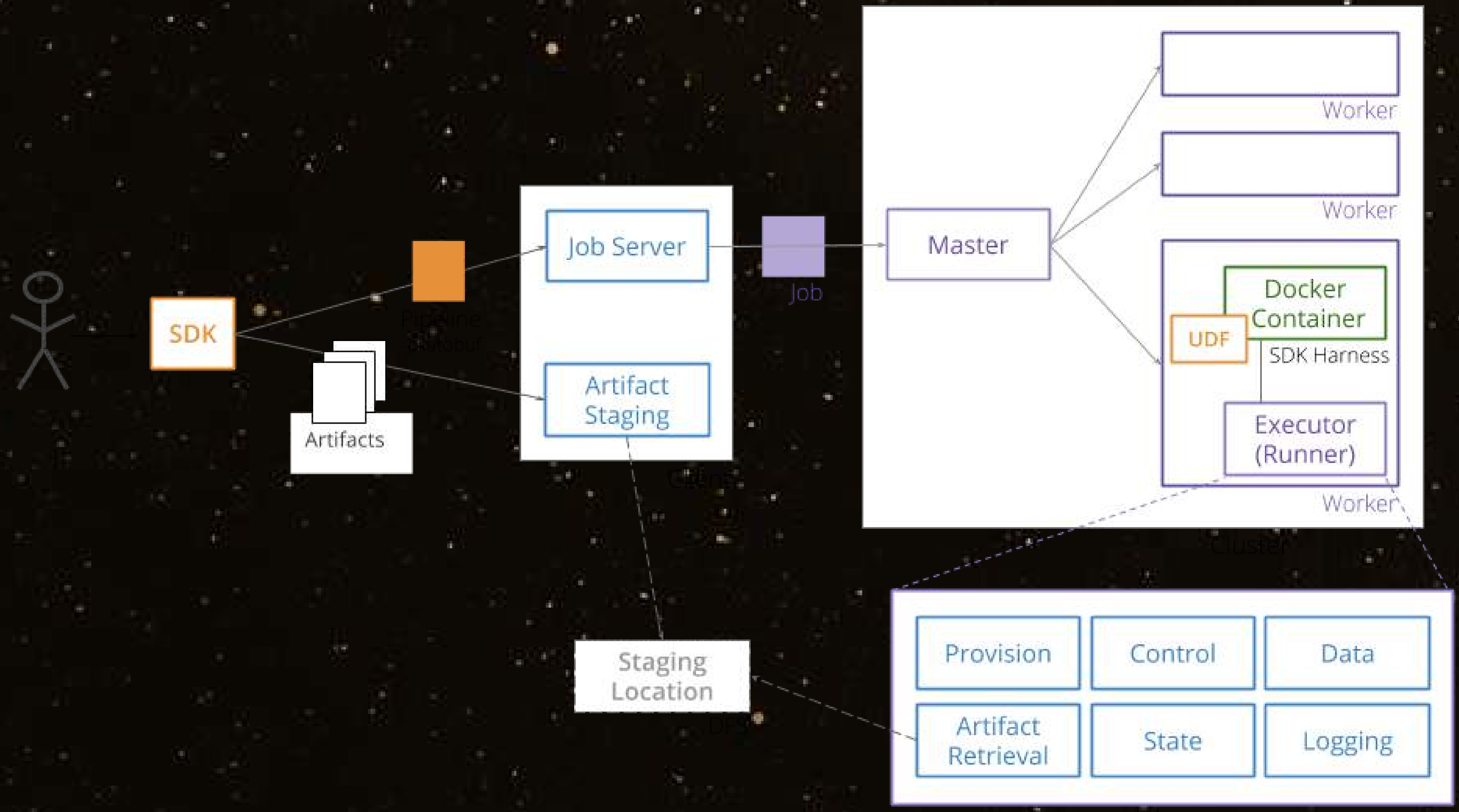
- Portierbar: Code abstraction that can be executed on different backend runners
- Vereinheitlicht: Unified Batch and Streaming API
- Erweiterbare Modelle und SDK: Extensible API to define custom sinks and sources

The Apache Beam Vision

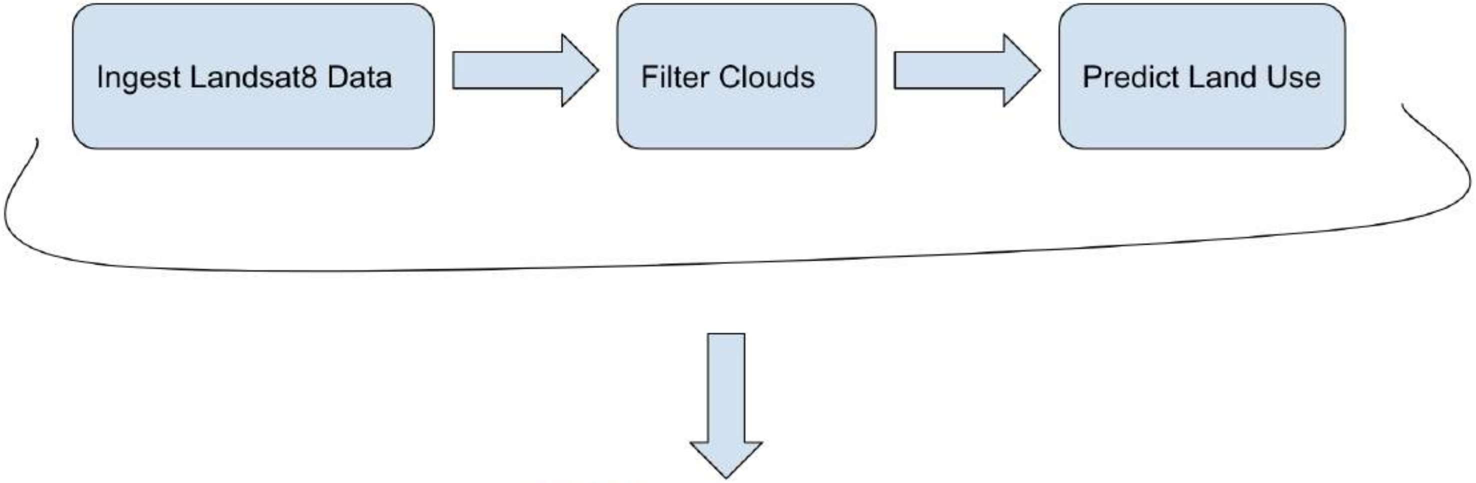
- End Users: Create pipelines in a familiar language
- SDK Writers: Make Beam concepts available in new languages
- Runner Writers: Support Beam pipelines in distributed processing environments



Portable Beam Architecture Overview


SDK
Artifacts
Job Server
Artifact Staging
Job
Master
Worker
Worker
Docker Container
UDF
SDK Harness
Executor (Runner)
Worker
Staging Location
Provision
Control
Data
Artifact Retrieval
State
Logging

Inference Pipeline



Beam Inference Pipeline

```
pipeline_options = PipelineOptions(pipeline_args)
pipeline_options.view_as(SetupOptions).save_main_session = True
pipeline_options.view_as(StandardOptions).streaming = True

with beam.Pipeline(options=pipeline_options) as p:
    filtered_images = (p | "Read Images" >> beam.Create(glob.glob
    | "Batch elements" >> beam.BatchElements(0, known_args.batchs
    | "Apply Cloud Mask" >> beam.ParDo(CloudMaskFn.CloudMaskFn())

filtered_images | "Predict Land Use" >>
            beam.ParDo(Prediction.PredictionFn(known_args.models,
```

Conclusion

Paradise, CA over 2018 had a consistent Dry Vegetation patterns with NDVI < 0

Healthy Green Vegetation in the range [0.3, 1.0)

Future Work

Classify Rock Formations

Using Shortwave Infrared images (2.107 - 2.294 nm)

Radiant Energy reflected/transmitted per unit time (Radiant Flux)

$$\Phi_e = \frac{\partial Q_e}{\partial t}$$

Eg: Plants don't grow on rocks

https://en.wikipedia.org/wiki/Radiant_flux

Use images from Red band

Identify borders, regions without much details with naked eye - Wonder Why?

Images are in Red band

Unsupervised Learning - Clustering

Credits

- Jörn Kottmann (Apache OpenNLP)
- Jose Contreras, Kellen Sunderland (Amazon - Berlin)
- Apache Beam: Pablo Estrada, Ankur Goenka, Maximilian Michels (Google)

Links

- Earth on AWS: https://aws.amazon.com/earth/
- Apache Beam: https://beam.apache.org
- Apache Flink: https://flink.apache.org
- Slides: https://smarthi.github.io/Berlin-Buzzwords2019-Landsat

Fragen ???