

## -----Dataset 1: NREL Solar Power Data-----

Dataset 1 was obtained from NREL Website which provided us with the Solar Power data for various locations. The link to download this dataset is: <http://www.nrel.gov/grid/solar-power-data.html>. This data is hosted and processed on Dumbo.

Code:

Combiner.py: Code to combine all the files into a single file

Combiner.py-UPV\_DPV.py: Code to combine UPV and DPV data separately.

Solar.zip: MapReduce code to profile data and perform analytic on data

SolarCodeUPV\_DPV.zip: MapReduce code to extract monthly power generated by UPV and DPV units.

Snapshot for MapReduce on UPV:

```
rsk430@login-1-1:~/project
WROONG_REDUCE=0
File Input Format Counters
  Bytes Read=8475137516
File Output Format Counters
  Bytes Written=1009816
[rsk430@login-1-1:~/project]$ hadoop com.bd.solar.Solar /user/rsk430/project/comb_data_DPV.csv /user/rsk430/project/output
DPV1
16/12/12 18:19:39 INFO client.RMProxy: Connecting to ResourceManager at babar.es.its.nyu.edu/128.122.215.50:8032
16/12/12 18:19:39 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool
ol interface and execute your application with ToolRunner to remedy this.
16/12/12 18:19:40 INFO input.FileInputFormat: Total input paths to process : 1
16/12/12 18:19:40 INFO mapreduce.JobSubmitter: number of splits:69
16/12/12 18:19:40 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1480358343656_19897
16/12/12 18:19:40 INFO impl.YarnClientImpl: Submitted application application_1480358343656_19897
16/12/12 18:19:40 INFO mapreduce.Job: The url to track the job: http://babar.es.its.nyu.edu:8088/proxy/application_14803
58343656_19897/
16/12/12 18:19:40 INFO mapreduce.Job: Running job: job_1480358343656_19897
16/12/12 18:19:44 INFO mapreduce.Job: Job job_1480358343656_19897 running in uber mode : false
16/12/12 18:19:44 INFO mapreduce.Job: map 0% reduce 0%
16/12/12 18:19:53 INFO mapreduce.Job: map 1% reduce 0%
16/12/12 18:19:54 INFO mapreduce.Job: map 65% reduce 0%
16/12/12 18:19:55 INFO mapreduce.Job: map 90% reduce 0%
16/12/12 18:19:56 INFO mapreduce.Job: map 98% reduce 0%
16/12/12 18:19:57 INFO mapreduce.Job: map 100% reduce 0%
16/12/12 18:20:02 INFO mapreduce.Job: map 100% reduce 13%
16/12/12 18:20:03 INFO mapreduce.Job: map 100% reduce 28%
16/12/12 18:20:04 INFO mapreduce.Job: map 100% reduce 38%
16/12/12 18:20:05 INFO mapreduce.Job: map 100% reduce 50%
16/12/12 18:20:06 INFO mapreduce.Job: map 100% reduce 93%
```

Snapshot for MapReduce on DPV:

```
rsk430@login-1-1:~/project
drwx-----  - rsk430 users          0 2016-12-06 21:35 /user/rsk430/project/outputUPV
[rsk430@login-1-1:~/project]$ export HADOOP_CLASSPATH=Solar.jar
[rsk430@login-1-1:~/project]$ hadoop com.bd.solar.Solar /user/rsk430/project/comb_data_UPV.csv /user/rsk430/project/output
UPV1
16/12/12 18:18:46 INFO client.RMProxy: Connecting to ResourceManager at babar.es.its.nyu.edu/128.122.215.50:8032
16/12/12 18:18:47 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool
ol interface and execute your application with ToolRunner to remedy this.
16/12/12 18:18:47 INFO input.FileInputFormat: Total input paths to process : 1
16/12/12 18:18:47 INFO mapreduce.JobSubmitter: number of splits:64
16/12/12 18:18:47 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1480358343656_19896
16/12/12 18:18:47 INFO impl.YarnClientImpl: Submitted application application_1480358343656_19896
16/12/12 18:18:47 INFO mapreduce.Job: The url to track the job: http://babar.es.its.nyu.edu:8088/proxy/application_14803
58343656_19896/
16/12/12 18:18:47 INFO mapreduce.Job: Running job: job_1480358343656_19896
16/12/12 18:18:52 INFO mapreduce.Job: Job job_1480358343656_19896 running in uber mode : false
16/12/12 18:18:52 INFO mapreduce.Job: map 0% reduce 0%
16/12/12 18:18:57 INFO mapreduce.Job: map 2% reduce 0%
16/12/12 18:19:02 INFO mapreduce.Job: map 72% reduce 0%
16/12/12 18:19:03 INFO mapreduce.Job: map 97% reduce 0%
16/12/12 18:19:04 INFO mapreduce.Job: map 100% reduce 0%
16/12/12 18:19:09 INFO mapreduce.Job: map 100% reduce 6%
16/12/12 18:19:10 INFO mapreduce.Job: map 100% reduce 25%
16/12/12 18:19:11 INFO mapreduce.Job: map 100% reduce 38%
16/12/12 18:19:12 INFO mapreduce.Job: map 100% reduce 50%
16/12/12 18:19:13 INFO mapreduce.Job: map 100% reduce 50%
16/12/12 18:19:14 INFO mapreduce.Job: map 100% reduce 94%
16/12/12 18:19:15 INFO mapreduce.Job: map 100% reduce 95%
16/12/12 18:19:16 INFO mapreduce.Job: map 100% reduce 99%
16/12/12 18:19:17 INFO mapreduce.Job: map 100% reduce 100%
```

## -----Dataset 2: NLCD Data-----

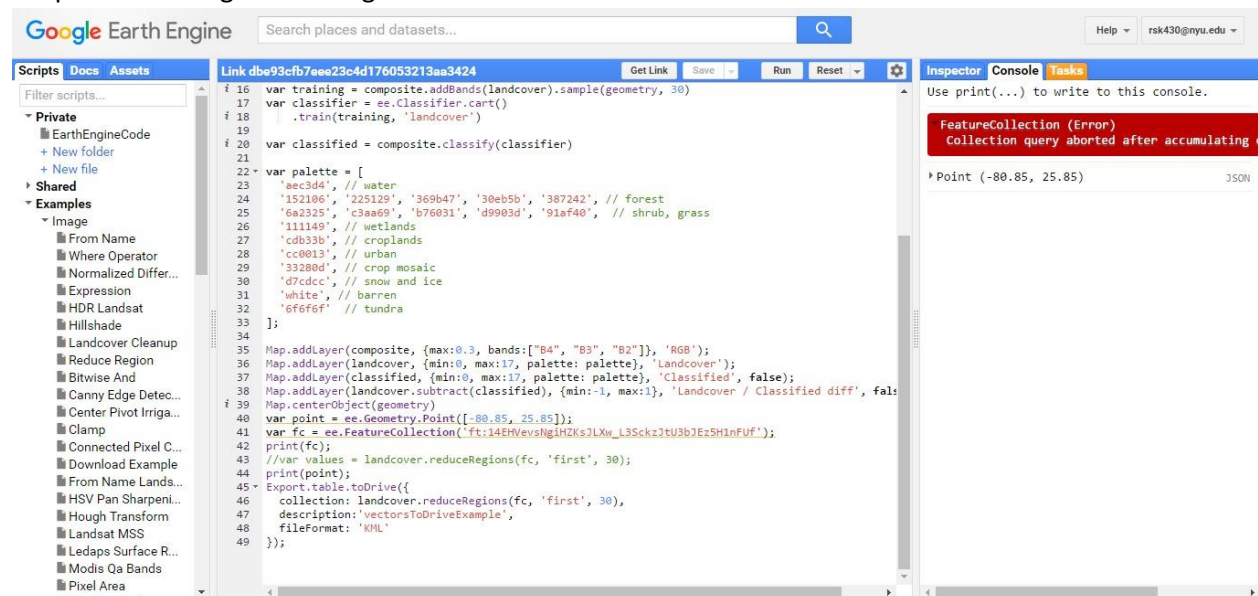
Dataset 2 was obtained from Multi-Resolution Land Characteristics (MRLC) Consortium website. As we were unable to process this data. We made use of Google earth Engine to extract the information from this data (Dataset was already hosted on Google Earth Engine). The link to download this dataset is: <http://www.mrlc.gov/nlcd2011.php>. But It's better to extract data directly from Google Earth Engine as special software (ArcGIS, ArcMap, FME workbench) are required to work with the data. This data was hosted and processed on Quickstart VM)

Google Earth Engine.txt: Code used to extract Landcover from NLCD data.

kmltoGeoJSON.py: Code to convert LandCover data from KML format to GeoJSON.

LandCodeforJSON.zip: MapReduce code to convert the GeoJSON data to CSV and isolate it in terms of Landcover so that it can be used in Impala.

Snapshot for Google Earth Engine:



---

## ANALYTICS

---

Impala Queries:

Queries to create tables in Impala for the Solar(UPV and DPV) and Landcover\_solar dataset

1. create table upv(latitude string, longitude string, month int, sum double) row format delimited fields terminated by ',' location '/user/cloudera/project/upv/';
2. create table dpv(latitude string, longitude string, month int, sum double) row format delimited fields terminated by ',' location '/user/cloudera/project/dpv/';
3. create table landsolar(longitude string, latitude string, landcover int) row format delimited fields terminated by ',' location '/user/cloudera/project/landcover\_solar/';

Query for Distributed PV (DPV) that can be used for solar power generation in houses and buildings and hence will include landcover like Urban, CropLands and Crop, Natural Veg. Mosaic.

4. `select dpv.latitude, dpv.longitude, dpv.month, dpv.sum from dpv, landsolar where CAST(dpv.latitude as double) = CAST(landsolar.latitude as double) and CAST(dpv.longitude as double) = CAST(landsolar.longitude as double) and landsolar.landcover in (12,13,14);`

Result of Analytic:

File	Edit	View	Search	Terminal	Help
40.55	-122.35	4	8770.040000000003		
40.65	-112.05	5	226882.90000000005		
41.05	-111.95	2	66521.100000000001		
41.15	-112.05	9	98258.999999999968		
41.25	-111.95	11	52746.200000000001		
41.35	-111.95	10	76365.499999999994		
42.35	-122.85	10	676.1089999999971		
42.95	-71.05	11	12638.600000000003		
43.55	-112.05	3	133.97100000000005		
43.95	-91.15	2	16308.400000000004		
44.05	-123.05	7	12882.780000000003		
44.55	-123.35	10	3004.9800000000002		
44.65	-93.05	10	26242.099999999998		
44.75	-91.25	11	21212.700000000001		
44.75	-93.05	11	17864.199999999999		
44.85	-123.05	10	3723.5100000000004		
44.95	-93.15	1	17791.100000000006		
44.95	-93.15	12	15038.300000000001		
44.95	-93.35	10	30123.799999999997		
47.05	-122.25	3	3937.1200000000012		
47.05	-122.35	2	2635.2000000000021		
47.15	-122.25	2	2796.6899999999999		
47.15	-122.35	1	1412.3299999999992		
47.15	-122.35	12	1624.1899999999995		
47.25	-122.35	11	1553.3299999999991		
47.25	-122.45	10	15937.4		
47.35	-122.25	11	2424.9999999999997		
47.45	-122.25	10	5928.3400000000016		
48.15	-122.15	2	5004.5900000000016		

+-----+-----+-----+-----+  
 Fetched 16644 row(s) in 4.06s  
 [quickstart.cloudera:21000] >

Query for Utility-scale PV (UPV) that can be used for Solar power farms, large space is required and hence will include Closed Shrubland, Savannas, Desert, Barren and Grasslands. The query for this is given below:

5. select upv.latitude, upv.longitude, upv.month, upv.sum from upv, landsolar where  
CAST(upv.latitude as double) = CAST(landsolar.latitude as double) and CAST(upv.longitude as double) = CAST(landsolar.longitude as double) and landsolar.landcover in (6,9,10,16);

Result for the analytic:

File Edit View Search Terminal Help				
43.25	-100.15	6	192749.19999999997	
43.25	-100.55	2	128220.60000000001	
43.25	-101.45	2	15487.900000000004	
43.25	-101.75	10	17672.299999999999	
43.25	-102.05	5	244945.20000000004	
43.25	-102.45	1	55285.400000000002	
43.25	-102.45	12	39444.300000000003	
43.25	-102.65	10	70311.999999999983	
43.35	-102.35	1	15816.1	
43.35	-102.35	12	11953.300000000001	
43.35	-102.45	11	91566.900000000011	
43.45	-102.35	11	26262.9	
43.45	-102.45	10	137940.9	
43.75	-100.05	2	60455.199999999992	
43.85	-101.15	10	52384.800000000003	
44.15	-121.35	5	51618.500000000011	
44.35	-101.15	3	48311.000000000001	
45.25	-104.15	11	80257.599999999995	
45.35	-104.15	10	176517.3	
46.75	-120.05	1	98976.300000000005	
46.75	-120.05	12	75070.499999999977	
46.75	-120.15	11	67593.499999999984	
46.75	-120.25	10	70816.700000000013	
46.85	-120.05	11	28541.599999999999	
46.85	-120.15	10	175182.30000000004	
47.35	-120.35	1	35758.199999999992	
47.35	-120.35	12	27738.199999999999	
47.45	-120.25	1	67589.399999999973	
47.45	-120.25	12	49670.199999999985	
+-----+-----+-----+-----+-----+				
Fetched 4152 row(s) in 1.60s				
[quickstart.cloudera:21000] > █				

Query to determine locations with highest UPV power generation potential (without combining with landcover data)

6. select latitude, longitude, sum(sum) as power from upv group by latitude, longitude order by power DESC limit 10;

Snapshot for the query:

```
[quickstart.cloudera:21000] > select latitude,longitude, sum(sum) as power from
upv group by latitude,longitude order by power DESC limit 10;
Query: select latitude,longitude, sum(sum) as power from upv group by latitude,l
ongitude order by power DESC limit 10
+-----+-----+-----+
| latitude | longitude | power |
+-----+-----+-----+
| 26.95    | -80.85    | 6239727.599999998 |
| 27.25    | -80.65    | 6030665.399999999 |
| 26.55    | -80.45    | 5954478.899999995 |
| 26.45    | -80.25    | 5935230.399999997 |
| 25.35    | -80.45    | 5903128.900000002 |
| 36.45    | -115.95   | 5601073.699999998 |
| 36.35    | -115.95   | 5564719.799999999 |
| 34.65    | -115.15   | 5464587.300000002 |
| 34.85    | -116.75   | 5429355.299999997 |
| 25.25    | -80.65    | 5323555.799999998 |
+-----+-----+-----+
Fetched 10 row(s) in 1.06s
[quickstart.cloudera:21000] > █
```

Result of Analytic:

latitude	longitude	power	Location
26.95	-80.85	6239727.6	FL, USA
27.25	-80.65	6030665.4	FL, USA
26.55	-80.45	5954478.9	FL, USA
26.45	-80.25	5935230.4	FL, USA
25.35	-80.45	5903128.9	FL, USA
36.45	-115.95	5601073.7	NV, USA
36.35	-115.95	5564719.8	NV, USA
34.65	-115.15	5464587.3	CA, USA
34.85	-116.75	5429355.3	CA, USA
36.95	-115.15	5158296.7	NV, USA
36.35	-119.95	4895116.1	NV, USA

Query to determine locations with highest UPV power generation potential (without combining with landcover data)

7. select latitude, longitude, sum(sum) as power from upv group by latitude, longitude order by power DESC limit 10;

Snapshot for the query:

```
[quickstart.cloudera:21000] > select latitude,longitude, sum(sum) as power from
dpv group by latitude,longitude order by power DESC limit 10;
Query: select latitude,longitude, sum(sum) as power from dpv group by latitude,l
ongitude order by power DESC limit 10
+-----+-----+-----+
| latitude | longitude | power |
+-----+-----+-----+
| 33.35    | -111.85   | 2340599.000000003 |
| 33.25    | -111.85   | 2331590 |
| 34.25    | -118.35   | 2263532.500000002 |
| 34.05    | -118.55   | 2263496.900000001 |
| 34.15    | -118.65   | 2254228.600000002 |
| 34.25    | -118.55   | 2250810.7 |
| 34.05    | -118.45   | 2249373.400000005 |
| 34.15    | -118.55   | 2245489.100000005 |
| 33.55    | -112.25   | 2243462.900000003 |
| 34.15    | -118.35   | 2233602.000000005 |
+-----+-----+-----+
Fetched 10 row(s) in 0.93s
[quickstart.cloudera:21000] > █
```

Result of analytic:

latitude	longitude	power	Location
33.35	-111.85	2340599	AZ, USA
33.25	-111.85	2331590	AZ, USA
34.25	-118.35	2263532.5	CA, USA
34.05	-118.55	2263496.9	CA, USA
34.15	-118.65	2254228.6	CA, USA
34.25	-118.55	2250810.7	CA, USA
34.05	-118.45	2249373.4	CA, USA
34.15	-118.55	2245489.1	CA, USA
33.55	-112.25	2243462.9	AZ, USA
34.15	-118.35	2233602	CA, USA