Impala Queries:

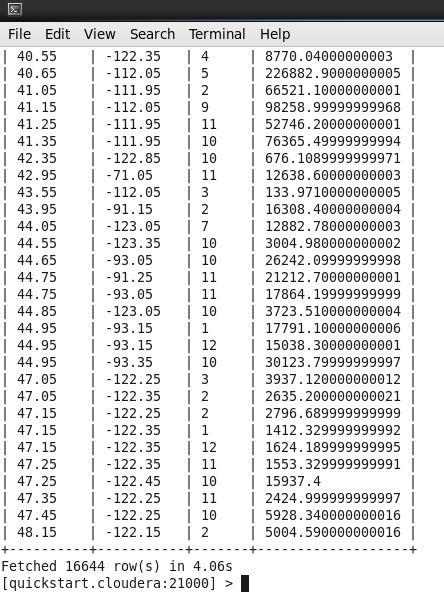
Queries to create tables in Impala for the Solar(UPV and DPV) and Landcover\_solar dataset

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.

1. 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);

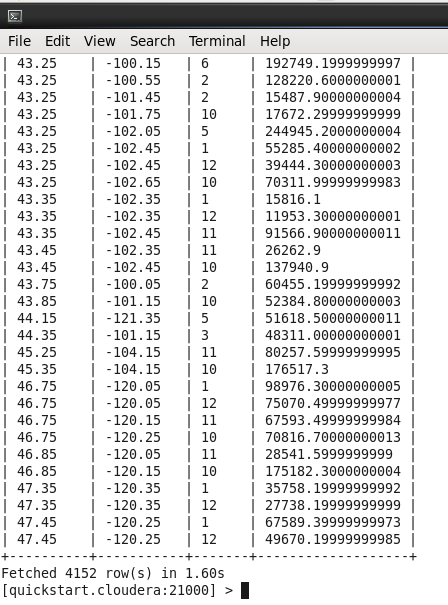
Result of Analytic:



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

1. 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);

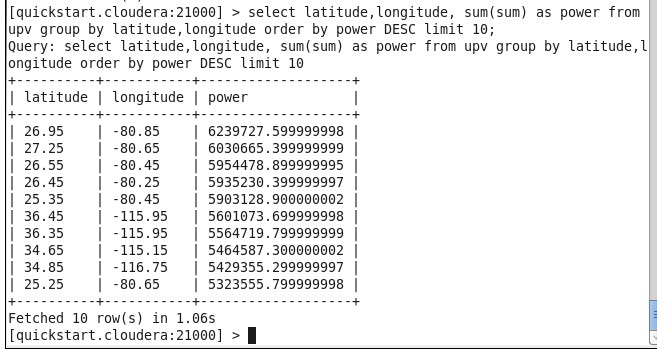
Result for the analytic:



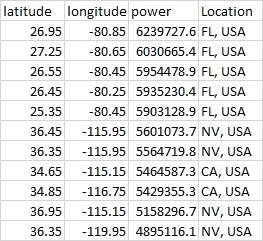
Query to determine locations with highest UPV power generation potential (without combining with landcover data)

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

Snapshot for the query:



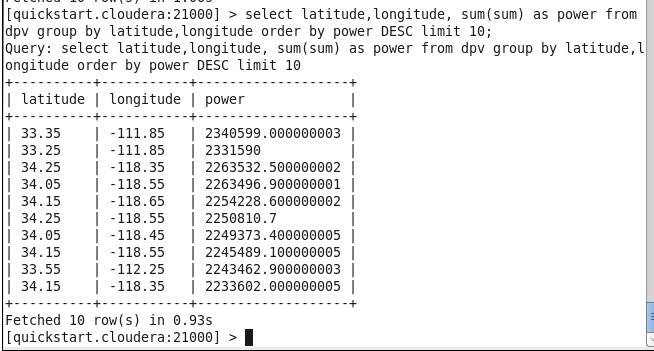
Result of Analytic:



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

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

Snapshot for the query:



Result of analytic:

