Question

Here, we have the temperatures collected every minute, from 20 top buildings all over the world. For this data analysis, you can download the necessary dataset from the below link https://github.com/prateekATacadgild/DatasetsForCognizant/blob/master/Sensor%20data- 20161205T052506Z.zip

In the above link there are two datasets; building.csv contains the details of the top 20 buildings all over the world and HVAC.csv contains the target temperature and the actual temperature along with the building Id.

HVAC (heating, ventilating/ventilation, and air conditioning) is the technology of indoor and vehicular environmental comfort. Its goal is to provide thermal comfort and acceptable indoor air quality.

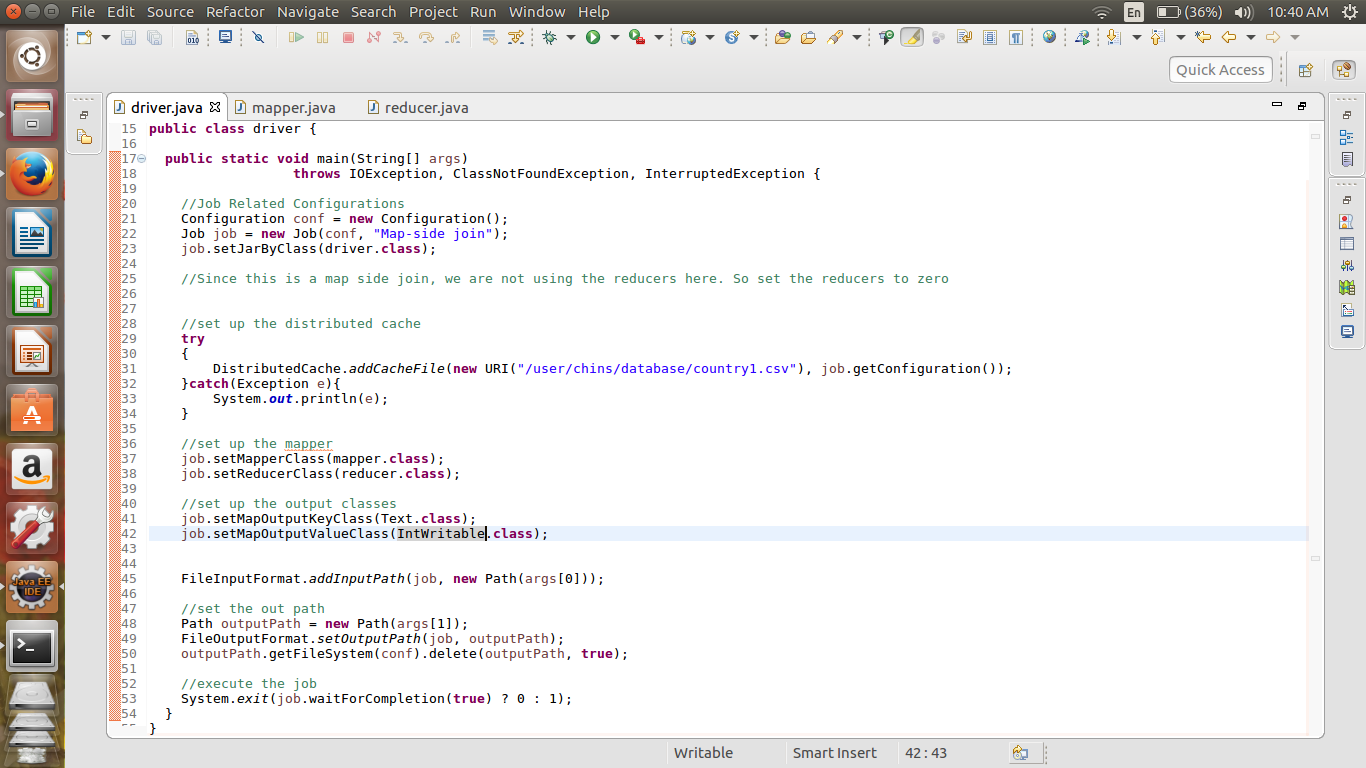
Through the HVAC sensors, we will get the temperature of the buildings. Here are the columns that are present in the datasets:

Building.csv – BuildingID, BuildingMgr, BuildingAge, HVACproduct,Country

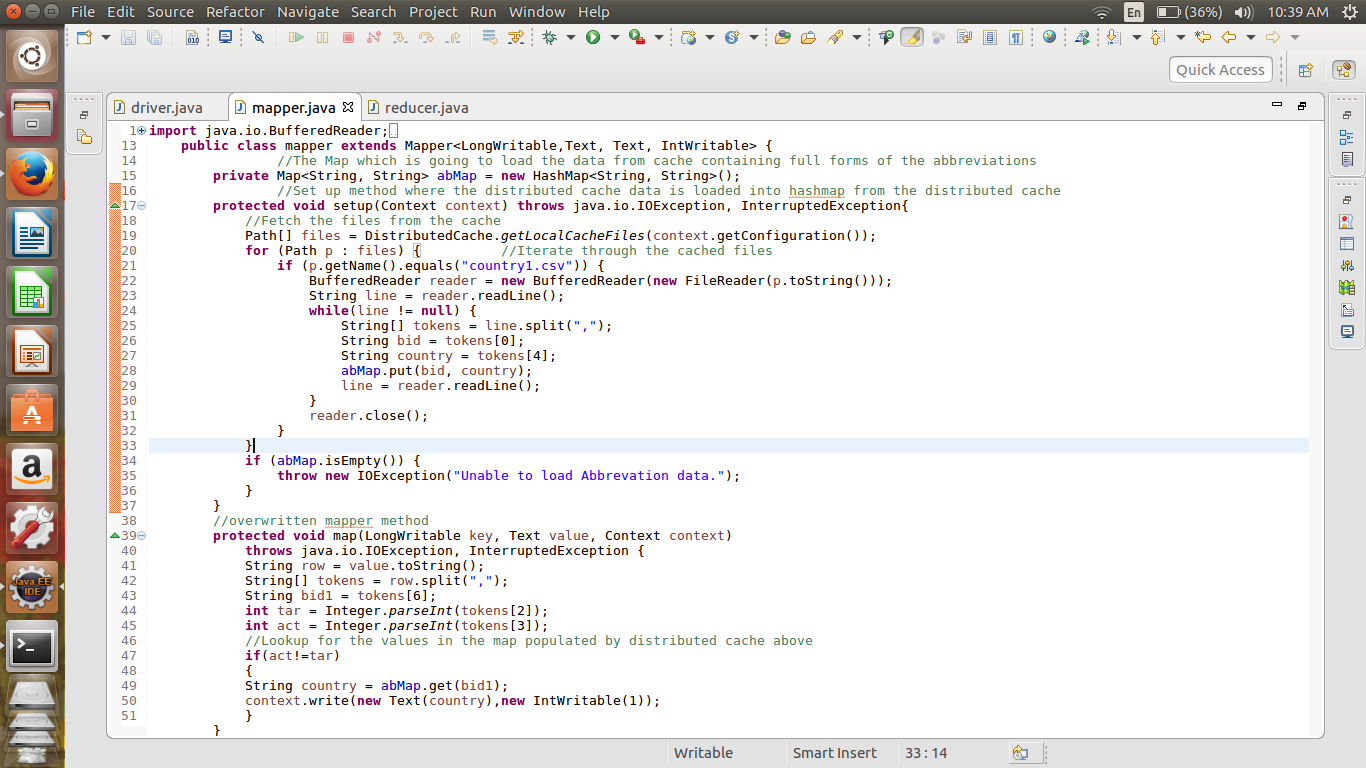
HVAC.csv – Date, Time, TargetTemp, ActualTemp, System, SystemAge, BuildingID

● Using Distributed cache join the two datasets and find out the number of temperature variations in each country using mapreduce.

//Driver

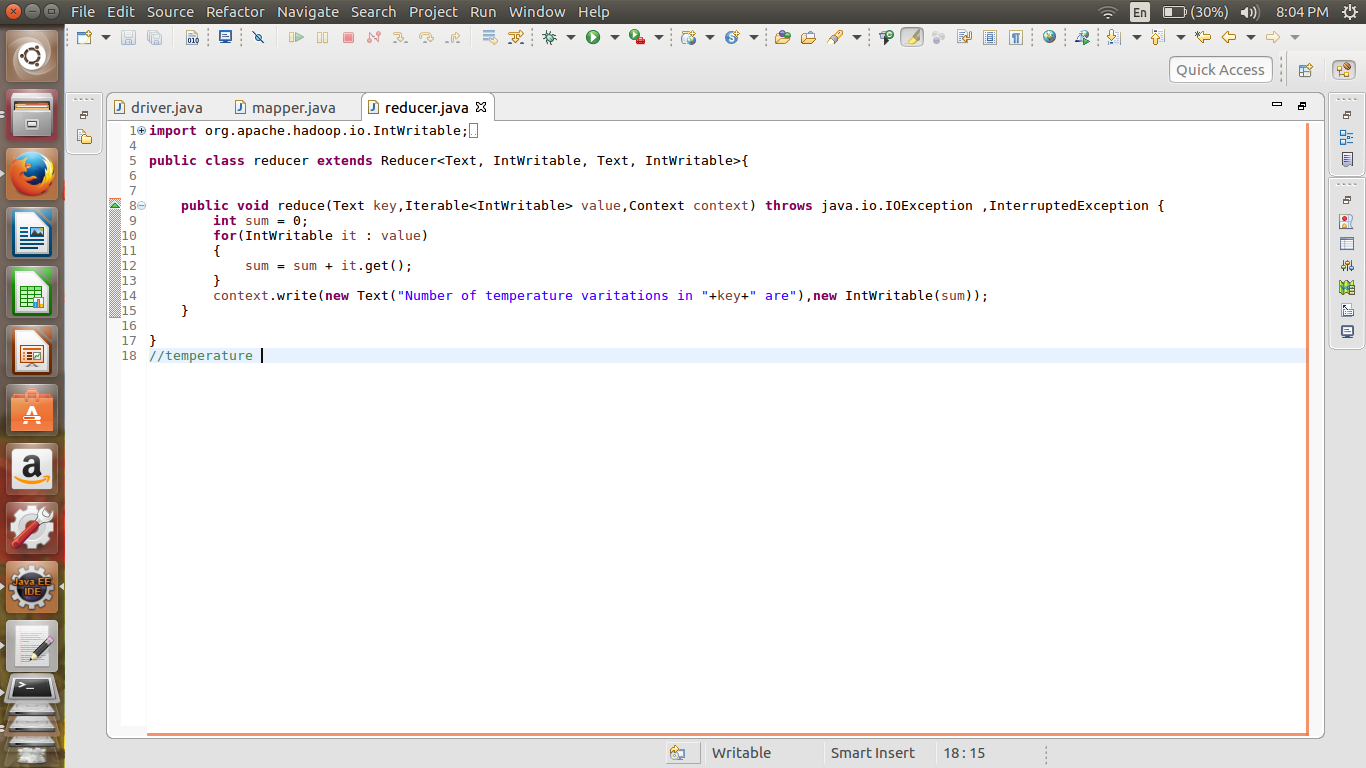


//Mapper Class

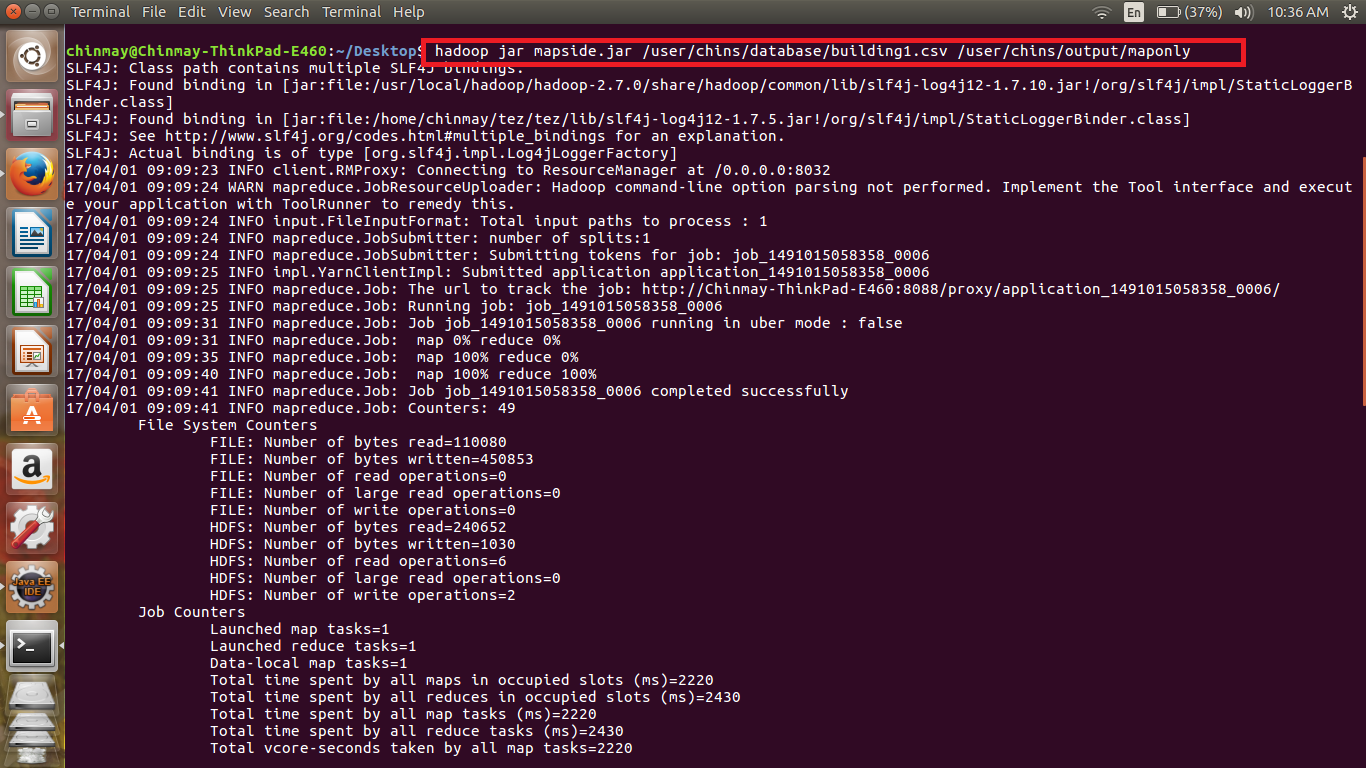


Here Using the Distributed cache for storing the file in cache for easy access. And performing the join in the mapper class.

//Reducer Class



//implementation Command



//Output

