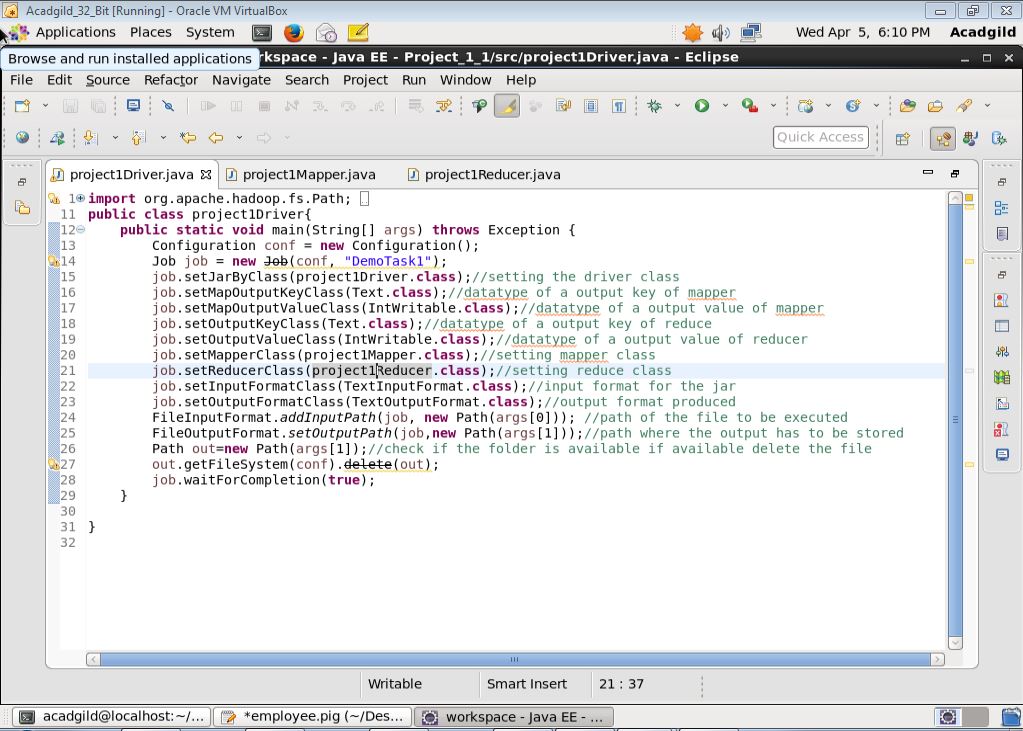
1. Write a mapreduce and pig program to calculate the number of cases investigated under each FBI code

Here we have a driver, mapper and reducer. As the driver is used to set the classes input and output formats and the configurations. The flow of the program is

Here the mapper is used to take the FBI Code as the key and the integer 1 as the value. The key is set as the code and the number of time id is invoked as many numbers of 1 are added in the value of the key.

Then the key value pair is passed to the reducer and the reducer is used to add or count the number of values to find the number of cases under each FBI code.

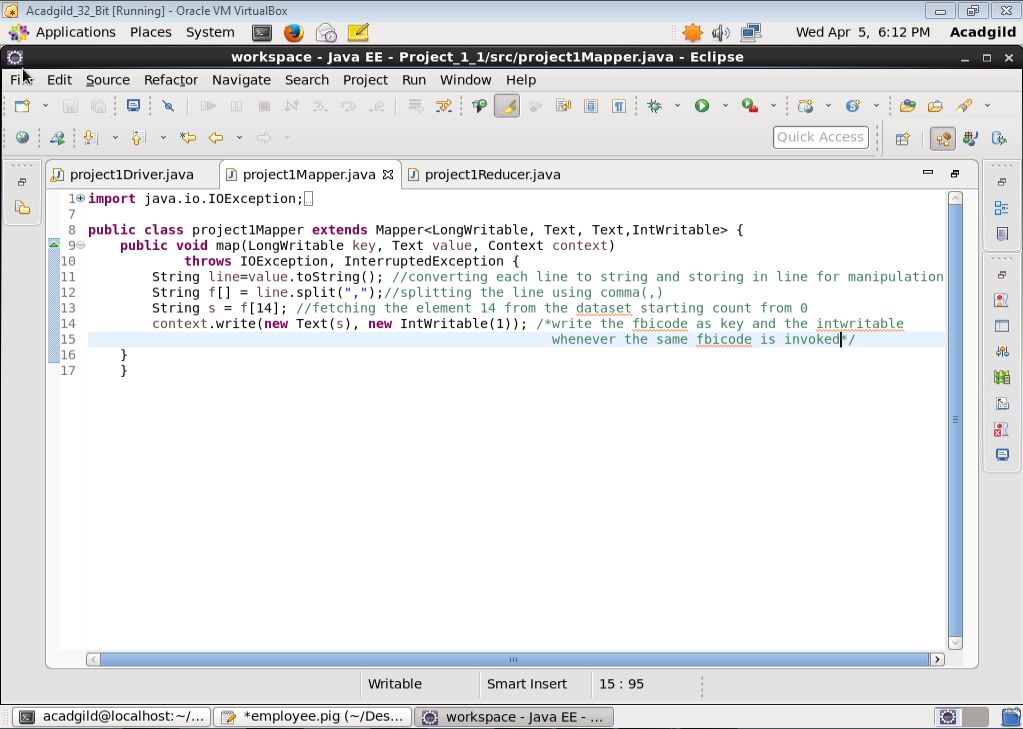
Here, driver class is used to set the mapper class, reducer class, number of reducers, file input format and the file output format. Also it can be used to check if the output path is available and if present delete the path and create it again. It is also used to wait till the job is being completed.



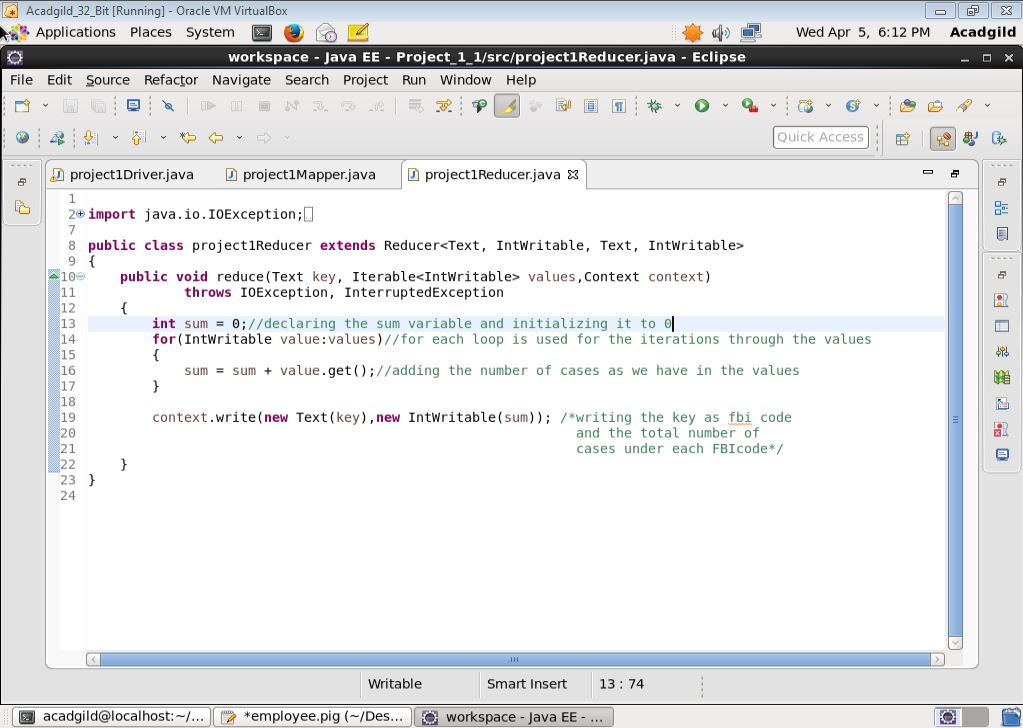
Here, in the mapper class we will process the data line by line and stored as a key value pair.

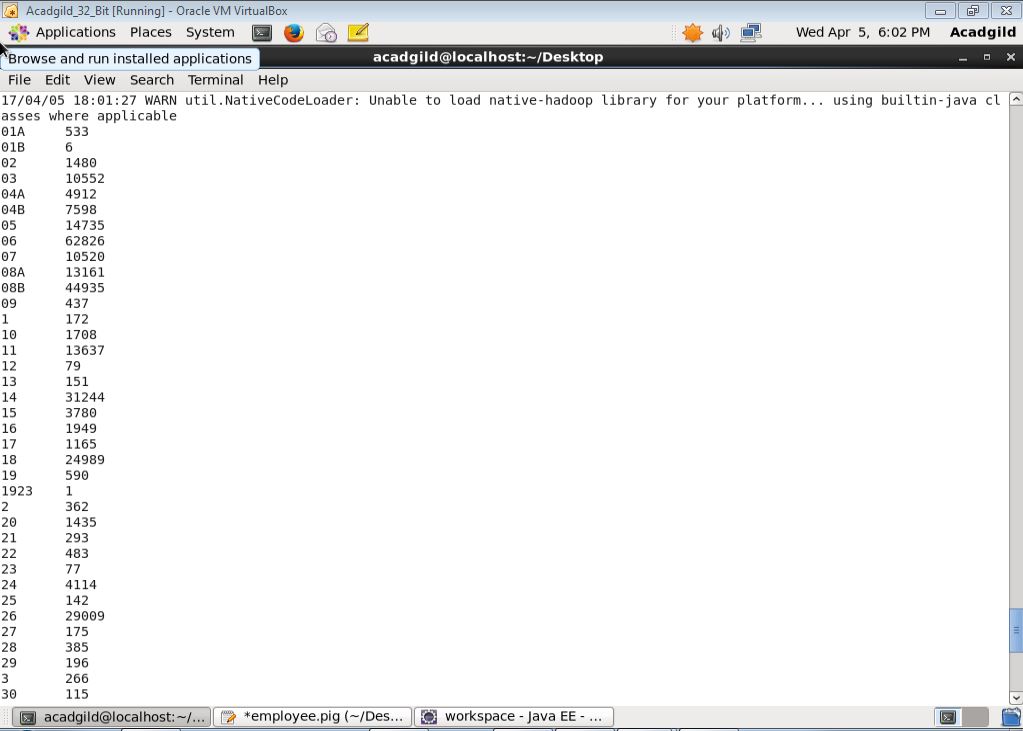
If we want some conditions to be executed for the keys we can do it here. Output of the mapper is in key value pair. The default format of the input key value is the long writable and it is the bit off set of the starting character.

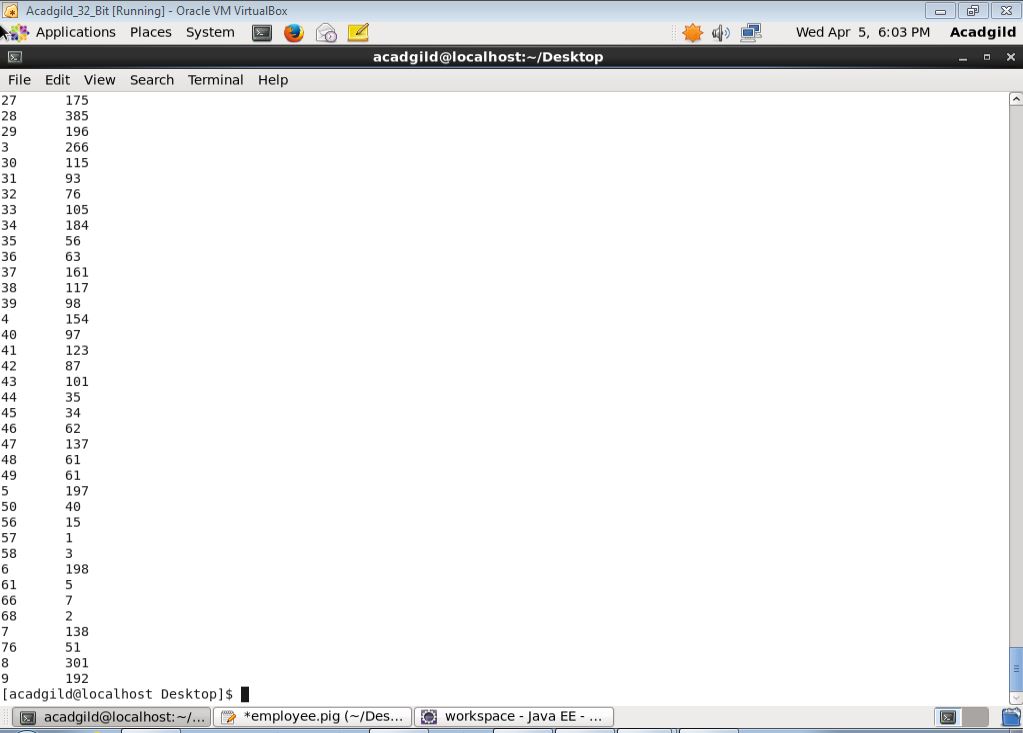
The line of the fetched is converted to the string and the string is split using comma and it is stored as an array and the required field is fetched and the key value pairs is being stored using context.



Shuffle and sort mechanism is done before it comes to the reducer and the data at the reducer is key and the values passed by the mapper. Then the reducer is used to derive the desired functionality from the data received and it performs the desired task. The conditions, manipulations are placed on the data here at reducer.







1. Write a mapreduce and pig program to calculate the number of cases investigated under FBI code 32.

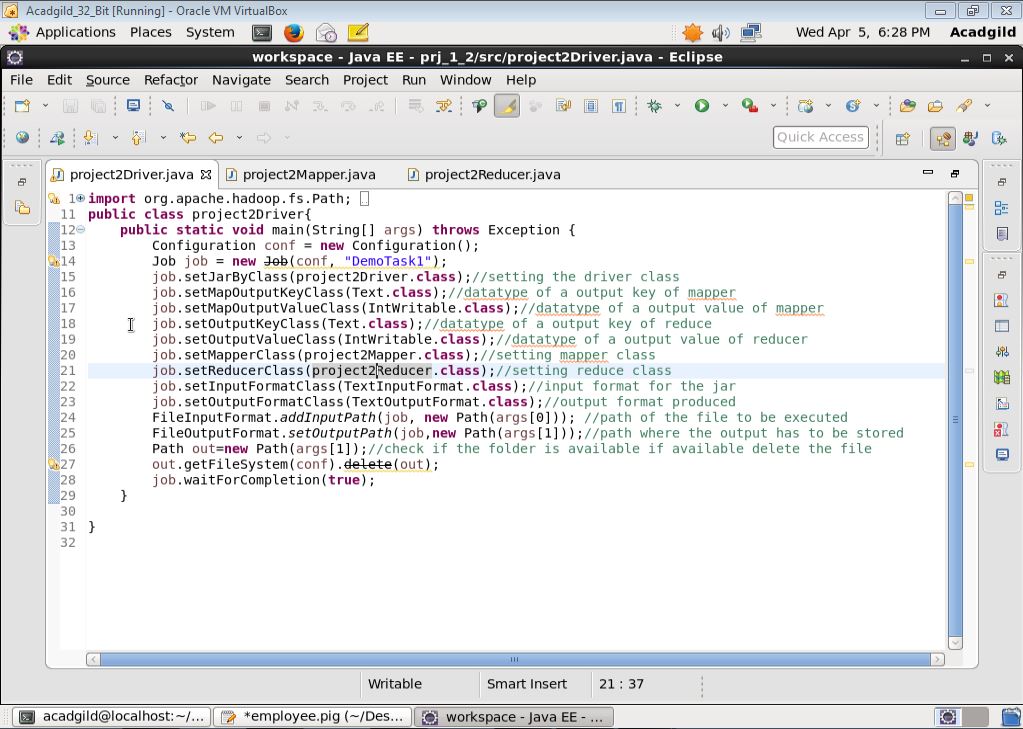
Here we have a driver, mapper and reducer. As the driver is used to set the classes input and output formats and the configurations. The flow of the program is

Here the mapper is used to take the FBI Code as the key and the integer 1 as the value. But we have placed a condition on key so that we can get only the case count for the FBI code 32. The number of time id is invoked as many numbers of 1 are added in the value of the key.

Then the key value pair is passed to the reducer and the reducer is used to add or count the number of values to find the number of cases under each FBI code. The condition that we placed in mapper for the code can be placed in the reducer. Iterable is used to iterate through the values and to add or count.

Then the output is stored as key value, which can be done according to our choices.

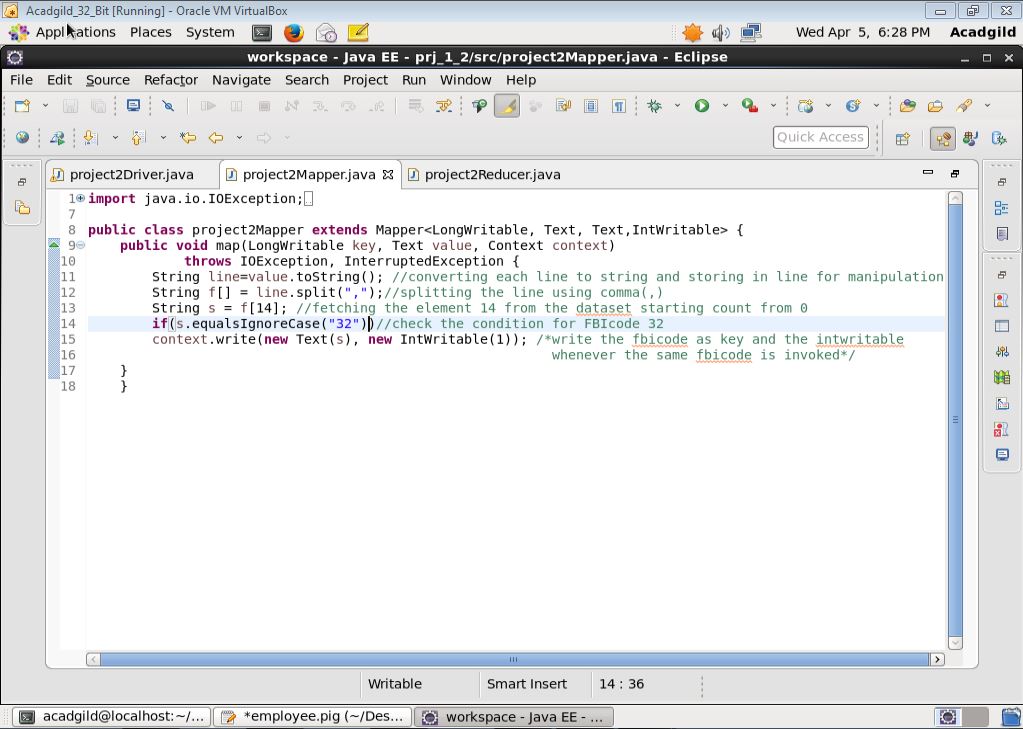
Here, driver class is used to set the mapper class, reducer class, number of reducers, file input format and the file output format. Also it can be used to check if the output path is available and if present delete the path and create it again. It is also used to wait till the job is being completed.



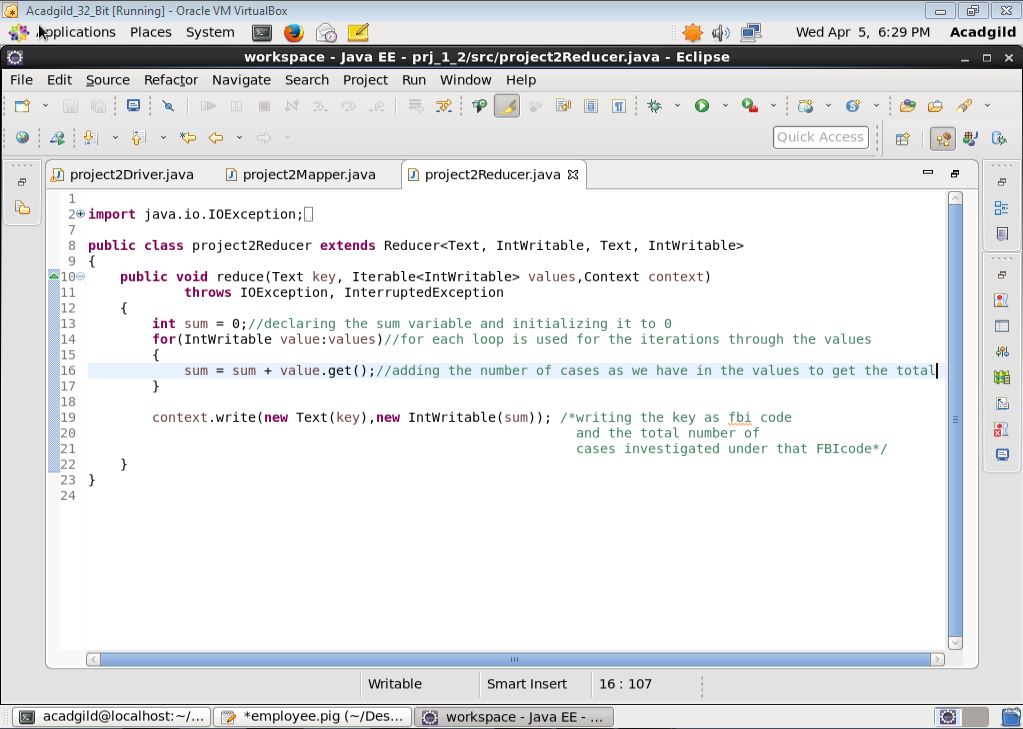
Here, in the mapper class we will process the data line by line and stored as a key value pair.

If we want some conditions to be executed for the keys we can do it here. Output of the mapper is in key value pair. The default format of the input key value is the long writable and it is the bit off set of the starting character.

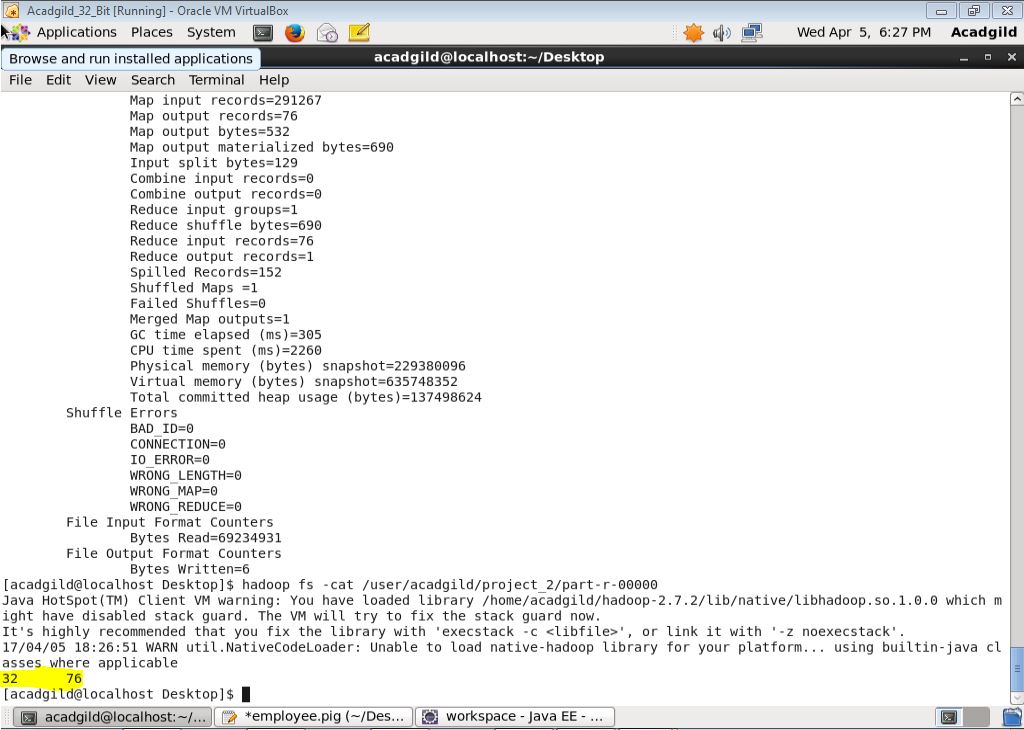
The line of the fetched is converted to the string and the string is split using comma and it is stored as an array and the required field is fetched and the key value pairs is being stored using context.



Shuffle and sort mechanism is done before it comes to the reducer and the data at the reducer is key and the values passed by the mapper. Then the reducer is used to derive the desired functionality from the data received and it performs the desired task. The conditions, manipulations are placed on the data here at reducer.



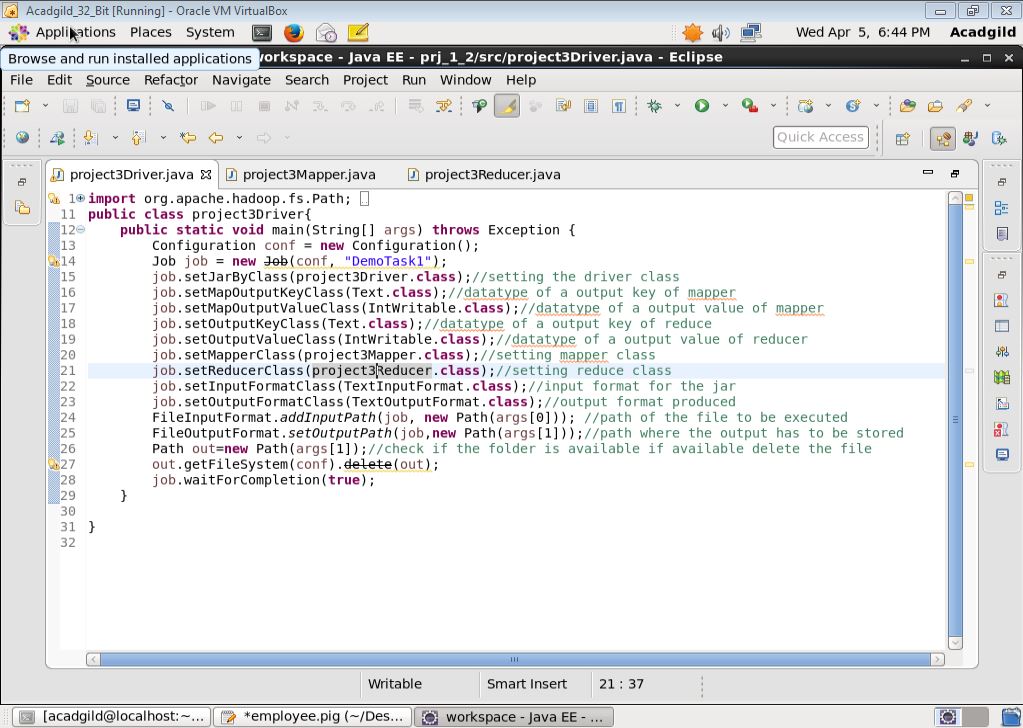
Here is the answer to the question. And the answer is 76.



1. Write a mapreduce and pig program to calculate the number of arrests in theft district wise.

Here in this program the number of thefts are calculated, all the thefts are considered here. So the matches condition is used to find all kind of the thefts. Also if the arrest is made or not the conditions for that has to be checked. So we have placed an if loop in the mapper to check the condition of the theft and the condition of the arrest to be true. Then the key value pairs is generated based on the district wise. So the key is district id and the values is the 1 as many time the district id is same when the conditions are satisfied. The key value pair is then passed to the reducer and at the reducer the iterable is utilized and the count of the arrests is generated and is stored in the key value pair.

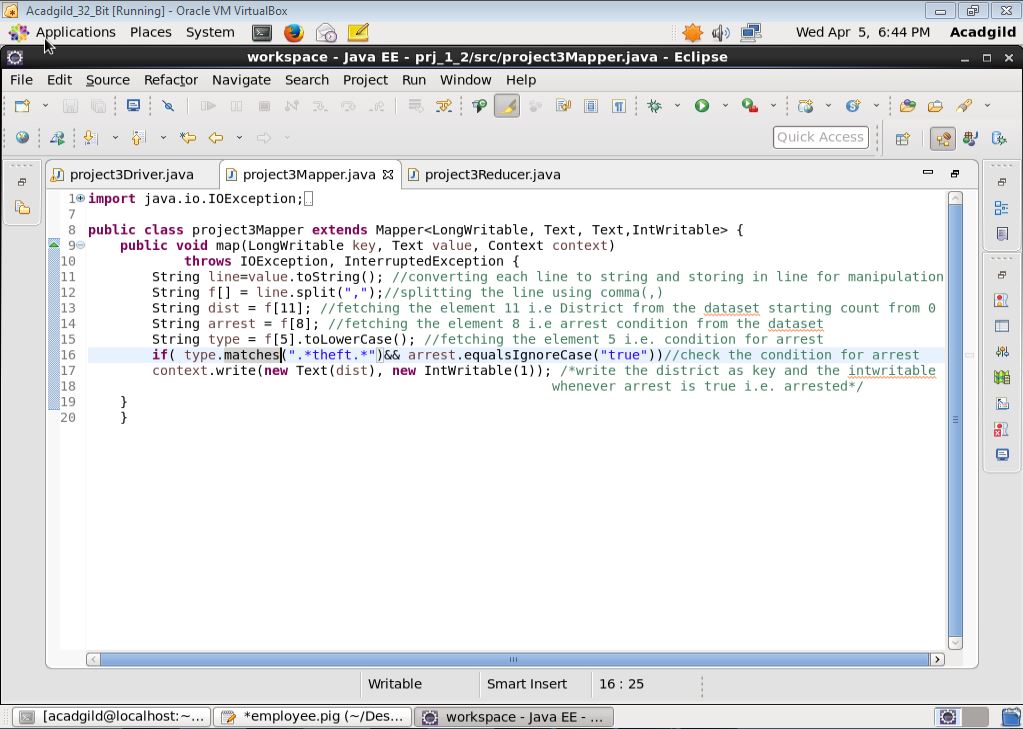
Here, driver class is used to set the mapper class, reducer class, number of reducers, file input format and the file output format. Also it can be used to check if the output path is available and if present delete the path and create it again. It is also used to wait till the job is being completed.



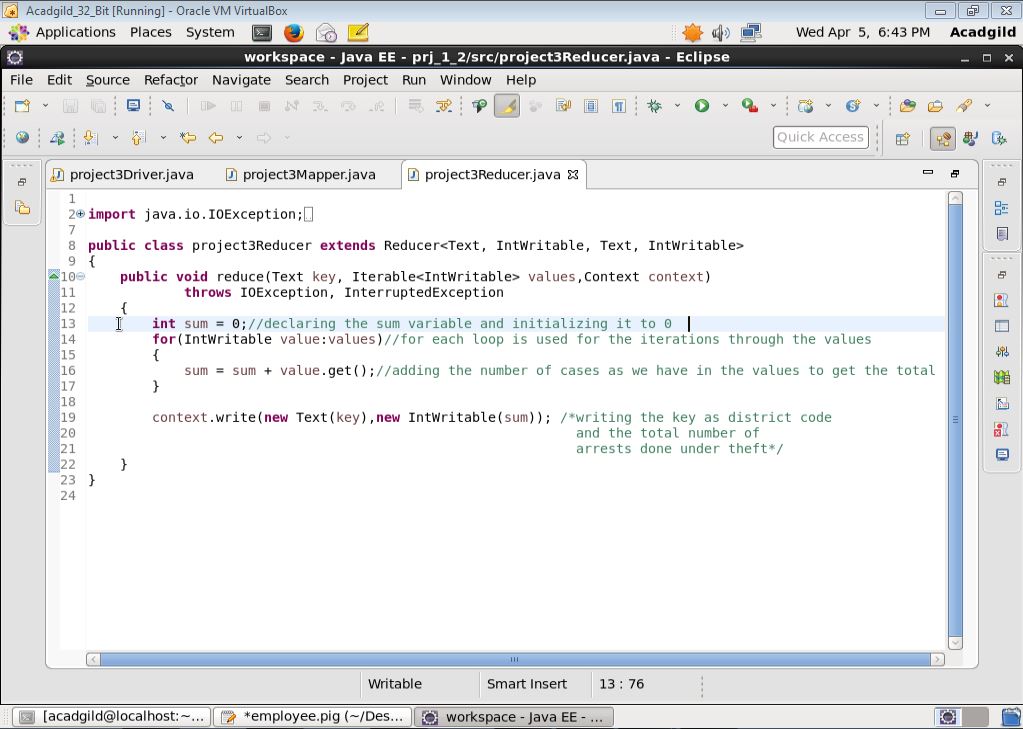
Here, in the mapper class we will process the data line by line and stored as a key value pair.

If we want some conditions to be executed for the keys we can do it here. Output of the mapper is in key value pair. The default format of the input key value is the long writable and it is the bit off set of the starting character.

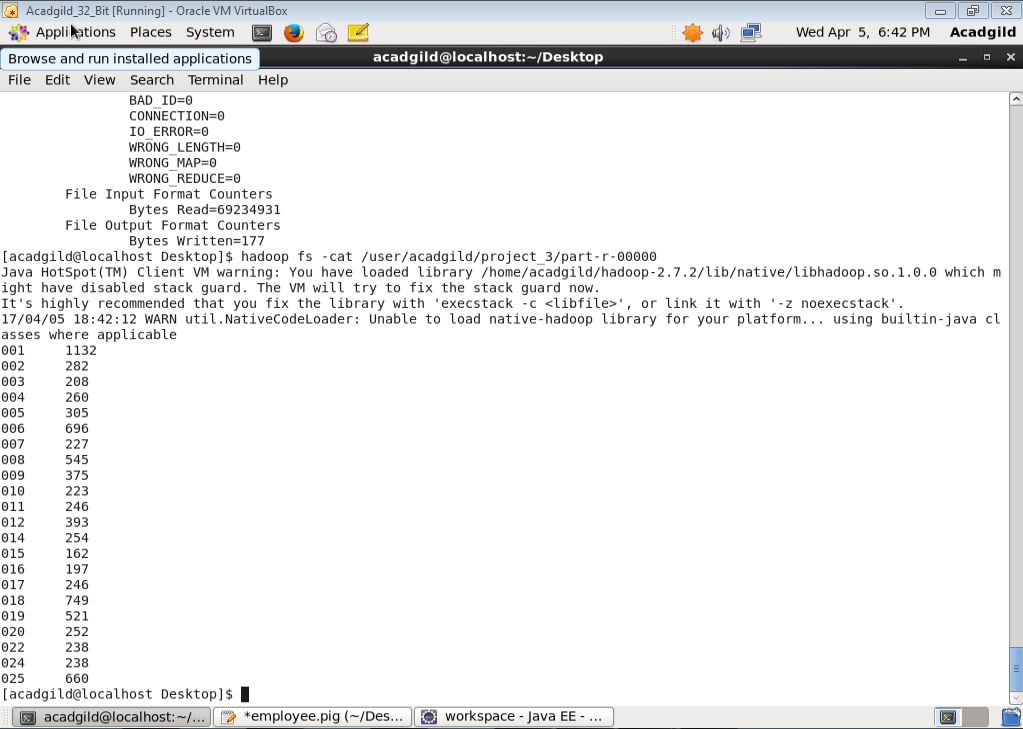
The line of the fetched is converted to the string and the string is split using comma and it is stored as an array and the required field is fetched and the key value pairs is being stored using context.



Shuffle and sort mechanism is done before it comes to the reducer and the data at the reducer is key and the values passed by the mapper. Then the reducer is used to derive the desired functionality from the data received and it performs the desired task. The conditions, manipulations are placed on the data here at reducer.



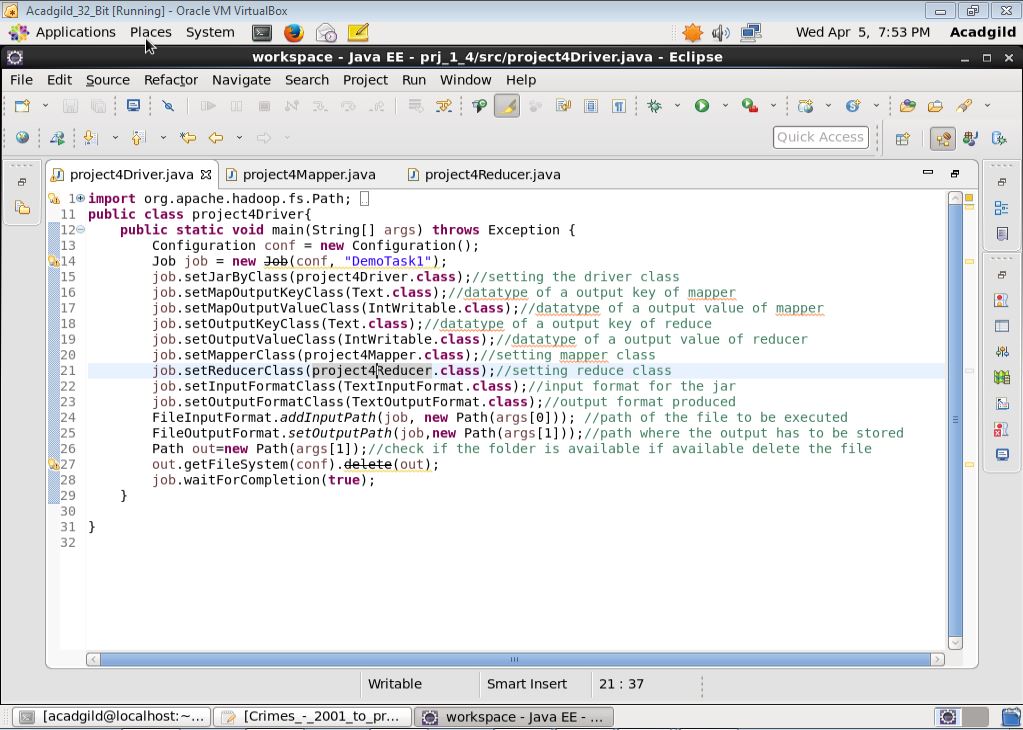
The answer for the required is as follows:



1. Write a mapreduce and pig program to calculate the number of arrests done between October 2014 and October 2015.

Here in this program the number of arrests during the specified time interval are calculated. We have included the start and end dates after 30,September,2014 11:59:59PM upto 1,November,2015 12:00:00AM. The duration is given from when we should start and when to stop the count. Here only one condition of arrest is to be checked. The condition of arrest has to be true. Here the key we have passed is number of arrest as we have to calculate all the arrest that took place in the given duration without any condition. The key value pair is then passed to the reducer and at the reducer the iterable is utilized and the count of the arrests is generated and is stored in the key value pair.

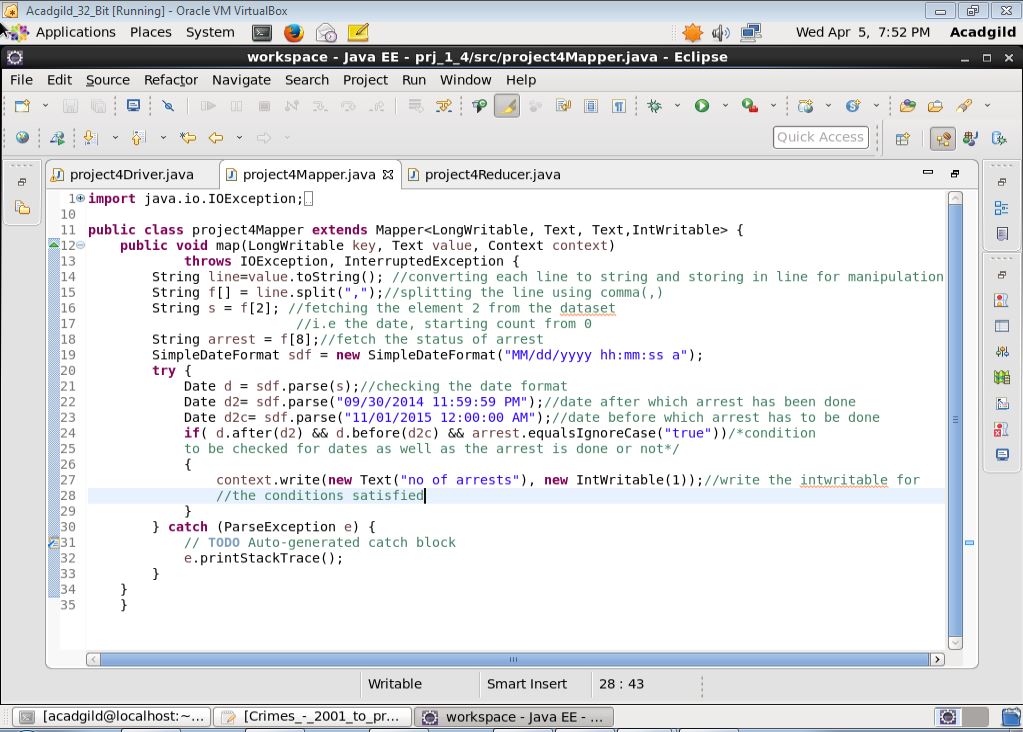
Here, driver class is used to set the mapper class, reducer class, number of reducers, file input format and the file output format. Also it can be used to check if the output path is available and if present delete the path and create it again. It is also used to wait till the job is being completed.



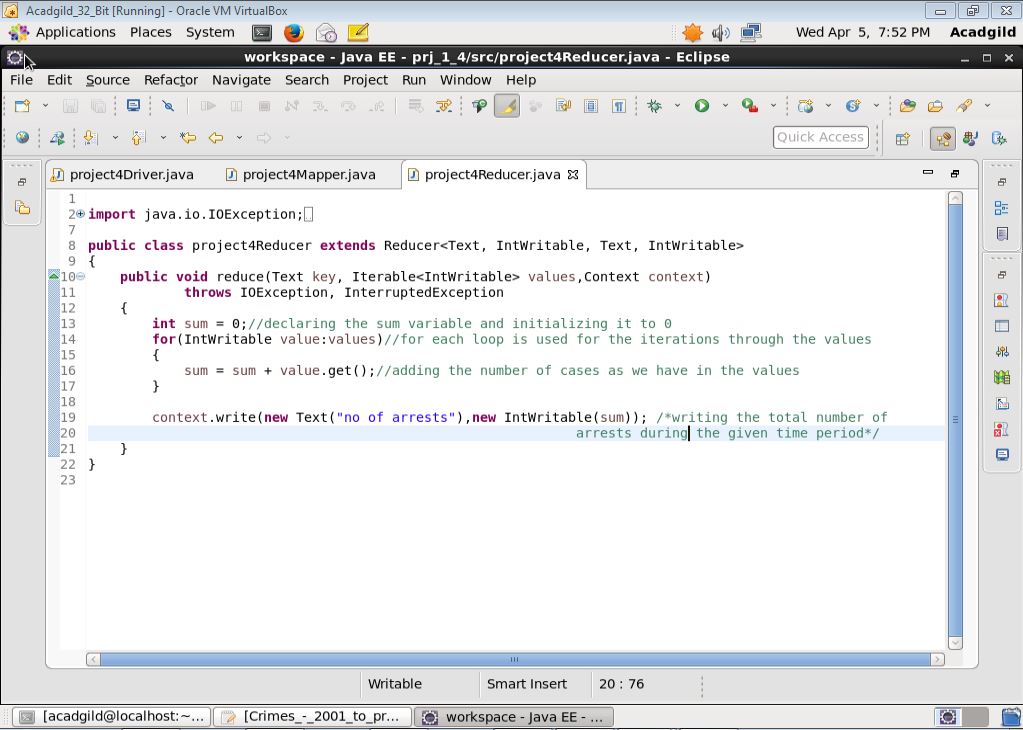
Here, in the mapper class we will process the data line by line and stored as a key value pair.

If we want some conditions to be executed for the keys we can do it here. Output of the mapper is in key value pair. The default format of the input key value is the long writable and it is the bit off set of the starting character.

The line of the fetched is converted to the string and the string is split using comma and it is stored as an array and the required field is fetched and the key value pairs is being stored using context.



Shuffle and sort mechanism is done before it comes to the reducer and the data at the reducer is key and the values passed by the mapper. Then the reducer is used to derive the desired functionality from the data received and it performs the desired task. The conditions, manipulations are placed on the data here at reducer.



The number of arrest are 63173

