11.6Find the top five kinds of Google Knowledge \_Occupation people who were guests in the show, in a particular time period i.e., between '1/11/99' and 6/11/99.

You need to write an UDF to compare the dates.

You can download the dataset from the below link

Daily Shows Guest Data

Dataset Description:

YEAR – The year the episode aired.

Google Knowledge Occupation -Their occupation or office, according to

Google’s Knowledge Graph. On the other hand, if they are not in there, how

Stewart introduced them on the program.

Show – Air date of the episode. Not unique, as some shows had more than

one guest

Group – A larger group designation for the occupation. For instance, U.S

Senators , U.S presidents, and former presidents are all under “politicians”

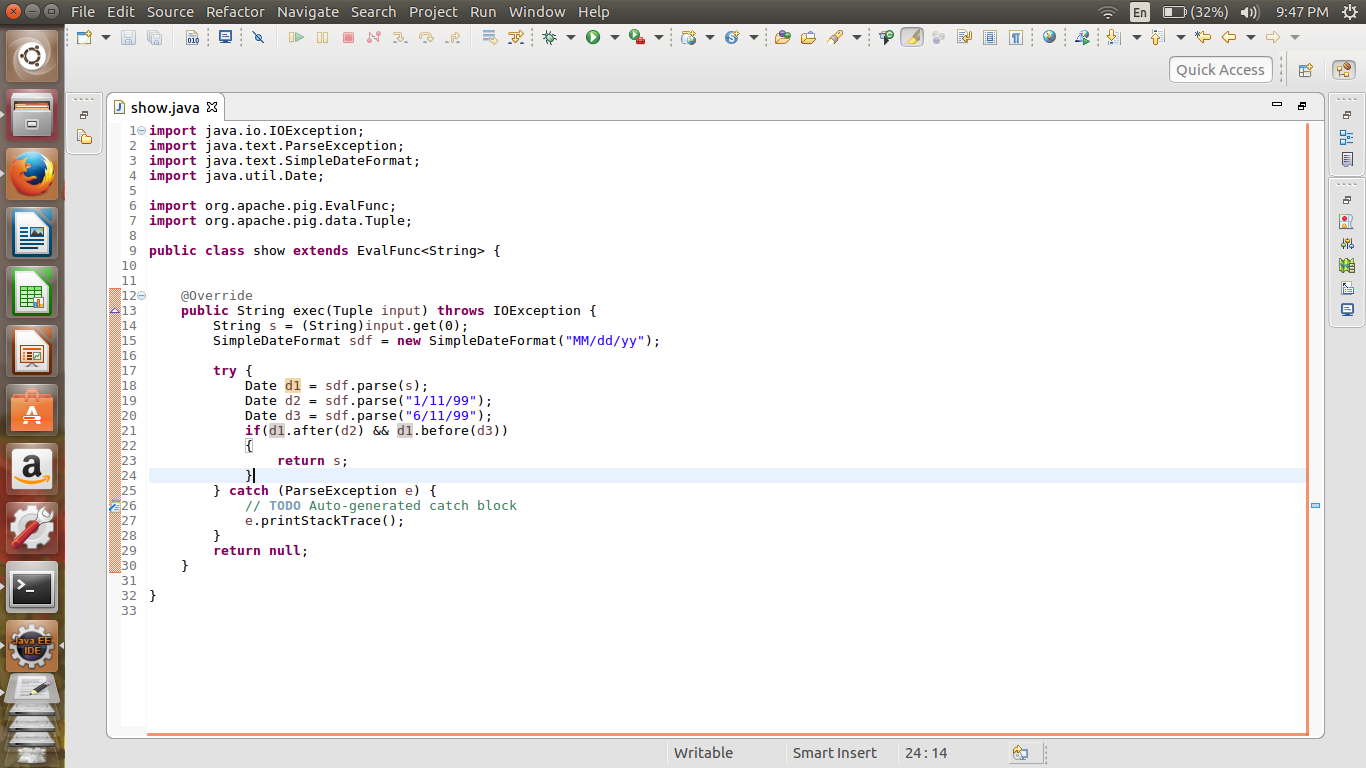
Raw\_ Guest \_List – The person or list of people who appeared on the show,

according to Wikipedia. The Google Knowledge \_Occupation only refers to one

of them in a given row.

**UDF :**

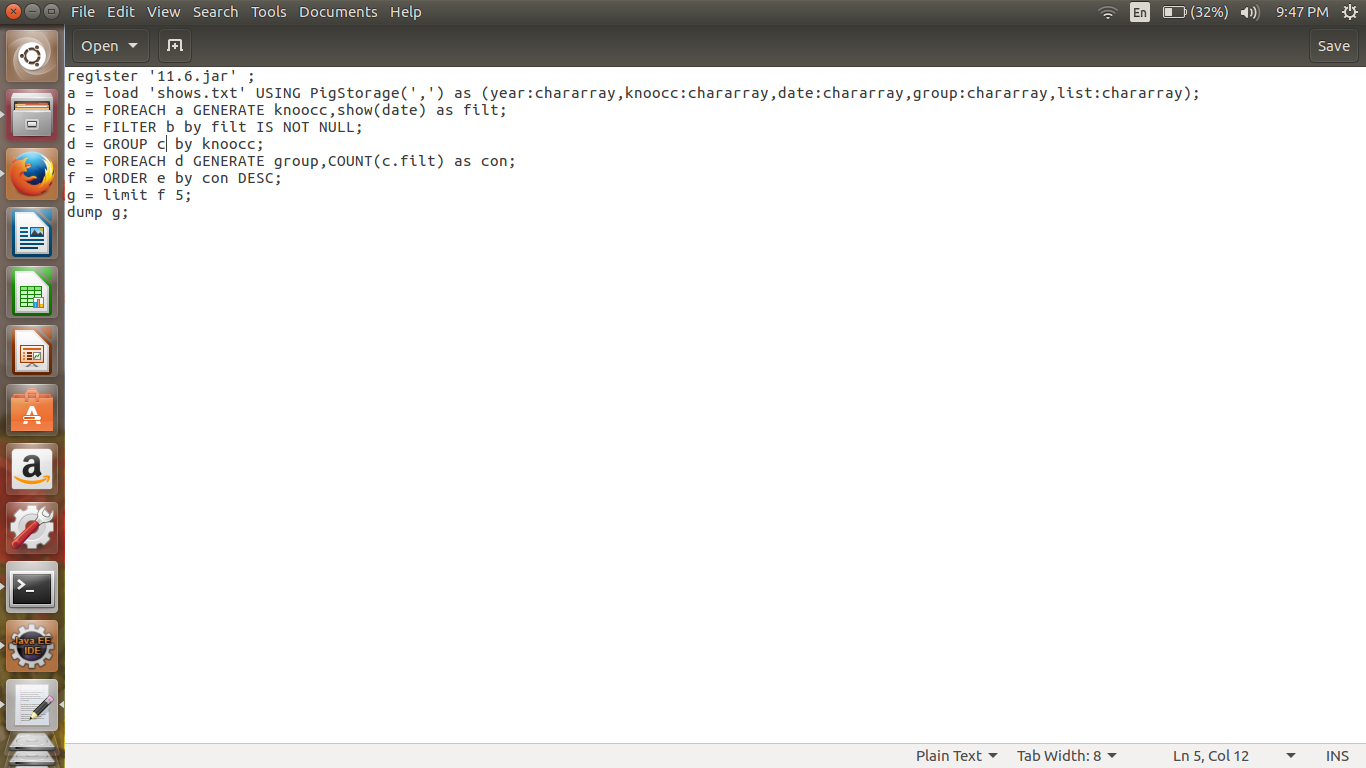
I’ve created a udf to filter the google knowledge occupation people who were in the show between 1/11/99 and 6/11/99.



d1.after(d2) && d1.before(d3) makes sure that the guests attended the show between1/11/99 and 6/11/99.

**Pig script:**

1. First I’ve registered the udf and loaded the dataset using register and load command respectively.
2. Next I’ve generated knowledge occupation and filtered it using the udf.
3. Next I’ve grouped knowledge occupation using group by command and counted it using count command.
4. At last I’ve ordered the list in descending order based on the count and limited to 5 by using limit command



**OUTPUT:**

