|  |
| --- |
| LOAD |
|  | This operator is used to load a file from a local system to a pig grunt shell. |
|  | example: A = LOAD '/home/acadgild/Student.txt' AS (id:int,points:int); |
|  | here in A a file Student.txt will be stored having two fields id and age. |
|  |  |
|  | STORE |
|  | When we need to store certain file from pig grunt shell to a local file system we used store command. |
|  | example: STORE A INTO 'outputfile' USING PigStorage('&'); |
|  | this will store all the tuples of A in outputfile by seperating it with \*. |
|  | later the file can be printed using CAT outputfile. |
|  |  |
|  | DUMP |
|  | To display all the tuples present in a file we use the DUMP command. |
|  | Example A has multiple tuples then to view the tuples we give: DUMP A; |
|  |  |
|  | FOREACH |
|  | FOREACH is used to execute certain operation on all the tuples of a dataset. |
|  | It generates data transformation based on certain condition of data. |
|  | it is written as: |
|  | X = FOREACH A GENERATE id\*10; |
|  |  |
|  | FILTER |
|  | It is used to select set of tuple based on certain condition. |
|  | A: |
|  | (1,2,7) |
|  | (4,2,1) |
|  | (4,3,7) |
|  |  |
|  | X = FILTER A BY f3==7; |
|  |  |
|  | then this will give: |
|  | X:(1,2,7) |
|  | (4,3,7) |
|  |  |
|  | GROUP BY |
|  | To group certain fields based on certain criterion group condition is used. |
|  | example: |
|  | X = GROUP A BY points; |
|  | this will group all the tuples according to their ages. |
|  |  |
|  | ORDER BY |
|  | This is used to sort the tuples based on asc or desc values of that field. |
|  |  |
|  | example: X = ORDER A BY A3 ASC; |
|  |  |
|  | DESCRIBE |
|  | To get the schema we use describe operator; |
|  | using DUMP we can simply get the tuples whereas using DESCRIBE we get the data types of all the fields, |
|  |  |
|  | A:(1,xyz,8); |
|  |  |
|  | DESCRIBE A; |
|  | This will give A:{id:int,name:chararray,value:int} |