|  |
| --- |
|  |
|  | **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 (name:chararray,age:int); |
|  | here in A a file Student.txt will be stored having two fields name 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 'myoutputfile' USING PigStorage('+'); |
|  | this will store all the tuples of A in myoutputfile by seperating it with +. |
|  | later the file can be printed using CAT myoutputfile. |
|  |  |
|  | **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 ceratin condition of data. |
|  | it is written as: |
|  | X = FOREACH A GENERATE age\*10; |
|  |  |
|  | **FILTER** |
|  | It is used to select set of tuple based on certain condition. |
|  | A: |
|  | (1,2,8) |
|  | (4,2,1) |
|  | (4,3,8) |
|  |  |
|  | X = FILTER A BY f3==1; |
|  |  |
|  | then this will give: |
|  | X:(4,2,1) |
|  |  |
|  |  |
|  | **GROUP BY** |
|  | To group certain fields based on certain criterion group condition is used. |
|  | example: |
|  | X = GROUP A BY age; |
|  | 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: U = ORDER A BY A3 DESC; |
|  |  |
|  | **DESCRIBE** |
|  | To get the schema we use describe oprator; |
|  | using DUMP we can simply get the tuples whereas using DESCRIBE we get the data types of all the fields, |
|  |  |
|  | A:(1,xyz,80); |
|  |  |
|  | DESCRIBE A; |
|  | This will give A:{id:int,name:chararray,marks:int} |