|  |
| --- |
|  |
| A. Load transaction records |
|  | txn = LOAD '/home/sowmya/Downloads/input files/txns1.txt' USING PigStorage(',') AS ( txnid, date, custid, amount:double, category, product, city, state, type); |
|  |  |
|  | B. Group transactions by customer |
|  |  |
|  | txnbycust = group txn by custid; |
|  | dump txnbycust; |
|  |  |
|  | C. Sum total amount spent by each customer |
|  |  |
|  | spendbycust = foreach txnbycust generate group as customer\_id, ROUND\_TO(SUM(txn.amount ),2) as totalsales; |
|  | dump spendbycust; |
|  |  |
|  | D. Order the customer records beginning from highest spender |
|  | custorder = order spendbycust by $1 desc; |
|  | dump custorder; |
|  |  |
|  | E. Select only top 10 customers |
|  | top10cust = limit custorder 10; |
|  | dump top10cust; |
|  |  |
|  | F. Join the transactions with customer details |
|  | top10join = join top10cust by $0, cust by $0; |
|  | describe top10join; |
|  | dump top10join; |
|  |  |
|  | G. Select the required fields from the join for final output |
|  | top10 = foreach top10join generate $0, $3, $4, $5, $6, $1; |
|  | dump top10; |
|  | top10order = order top10 by $5 desc; |
|  | describe top10order; |
|  |  |
|  | H. Dump and store the final output |
|  | dump top10order; |
|  | store top10order into '/home/hduser/pig\_result'; |
|  | \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* |
|  |  |
|  |  |
|  | CALCULATE TOTAL SALES |
|  | 1.groupbytype= group txn by type; |
|  | 2.describe groupbytype; |
|  | 3.totalbytype= foreach groupbytype GENERATE group as type,SUM(txn.amount)as sales; |
|  | 4.totalsale = group totalbytype all; |
|  | 5.totalsales = foreach totalsale GENERATE group as spendby,ROUND\_TO(SUM(totalbytype.sales),2) as total; |
|  | 6.describe totalsales; |
|  | totalsales: {total: double} |
|  | dump totalsales ; |
|  | 7. cashcredit = foreach totalbytype GENERATE $0,$1,($1/totalsales.total) as percent; |
|  | 8. dump cashcredit; |
|  | \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* |
|  |  |
|  |  |
|  | 1) groupbycustid = group cust by custid; |
|  | describe groupbycustid; |
|  | 2)totalsale= foreach groupbycustid GENERATE group as custid, SUM(txn.amount) as total; |
|  | 3)describe totalsale; |
|  | totalsale: {custid: bytearray,total: double} |
|  | 4)sale\_500plus = filter totalsale by total >500; |
|  | dump sale\_500plus; |
|  | 5) joincust = join cust by $0,sale\_500plus by $0; |
|  | 6)custsale = filter joincust by age <50; |
|  | describe custsale; |
|  | 7) dump custsale; |
|  |  |
|  |  |