**Pig commands**

data1 = load '/user/hadoop/AmazonSports.tsv' using PigStorage('\t') AS (marketplace, customer\_id, review\_id, product\_id, product\_parent, product\_title, product\_category, star\_rating, helpful\_votes, total\_votes, vine, verified\_purchase, review\_headline, review\_body, review\_date);

data = STREAM data1 THROUGH `tail -n +2` AS (marketplace, customer\_id, review\_id, product\_id, product\_parent, product\_title, product\_category, star\_rating, helpful\_votes, total\_votes, vine, verified\_purchase, review\_headline, review\_body, review\_date);

1. **Count of Daily Reviews**

dailydata = GROUP data by review\_date;

daily\_reviews = FOREACH dailydata GENERATE group as review\_date, COUNT(data.review\_id) as count;

orderdata = ORDER daily\_reviews BY count DESC;

store orderdata INTO '/pig1';

hadoop fs -ls /pig1

1. **Reviews per Rating**

product = GROUP data by star\_rating;

prod\_count = FOREACH product GENERATE group as star\_rating, COUNT (data.product\_id) as count;

store prod\_count INTO '/pig2';