

Homework – Do as much as you can...

- Count the number of missing value pattern in the Bosch manufacturing dataset
- Count unique number of users per track in last-fm dataset.
 - Input file:
<http://ocelma.net/MusicRecommendationDataset/lastfm-1K.html>
 - Unarchive: tar zxvf lastfm-dataset-1K.tar.gz
 - Using userid-timestamp-artid-artname-traid-traname.tsv
 - Get the number of unique users per track
- Implement Top-100 hash tags using tweets.txt
 - In this work, you should get top 100 hashtags in the tweets.txt.
Hint: You can create two MR jobs – one to get hash tag count and another to sort/write top 100 appearing hash tags

What to submit

- Report of explaining how the implementation works
- The source code of Map/Reduce
- Archive the files and submit it to cyber2010

Due

- Jan. 8th. 23:59