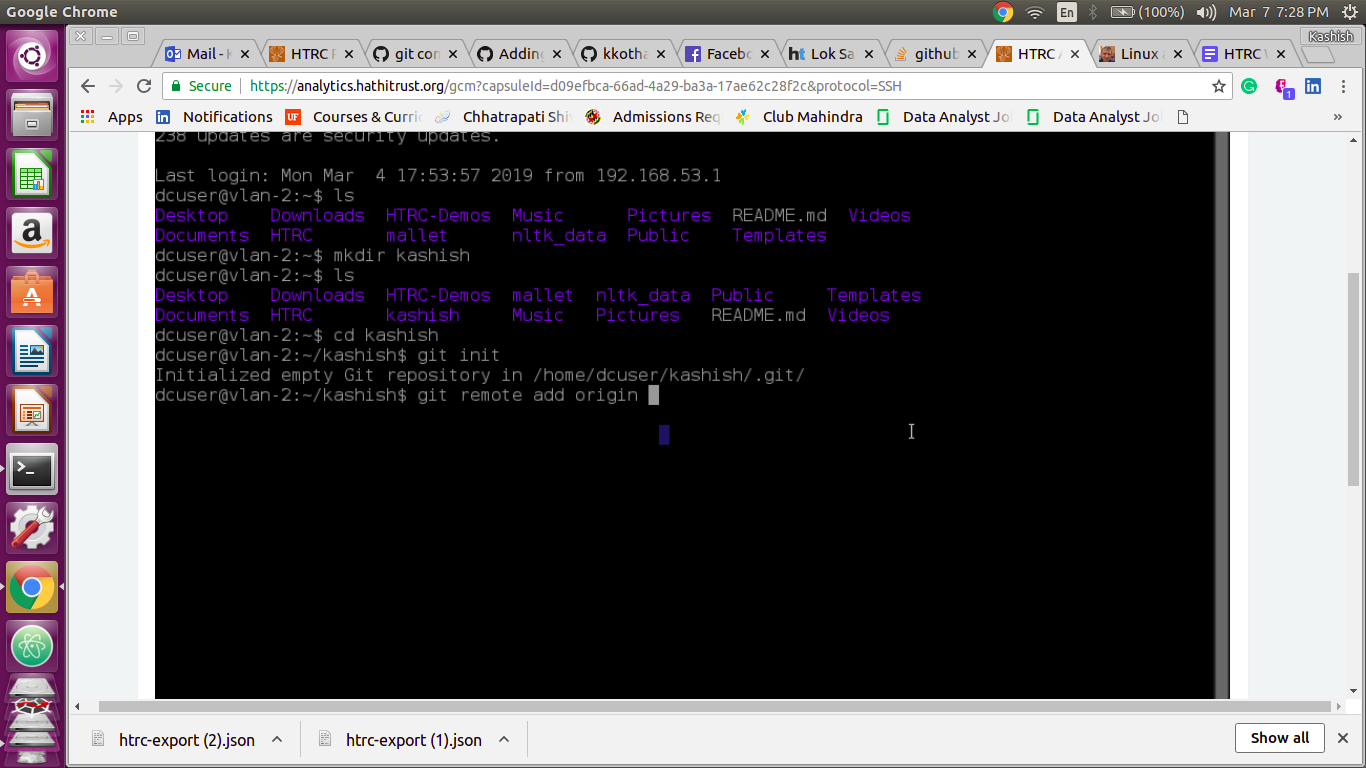
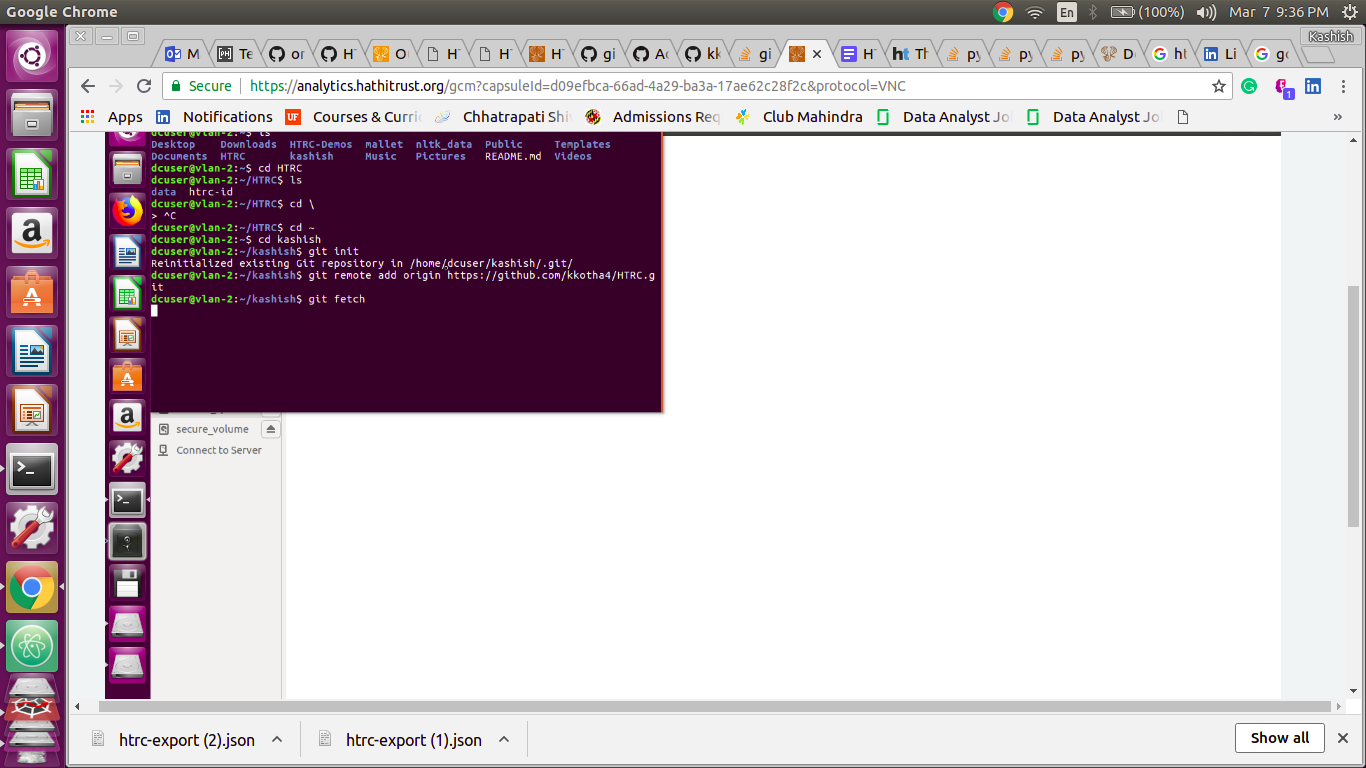
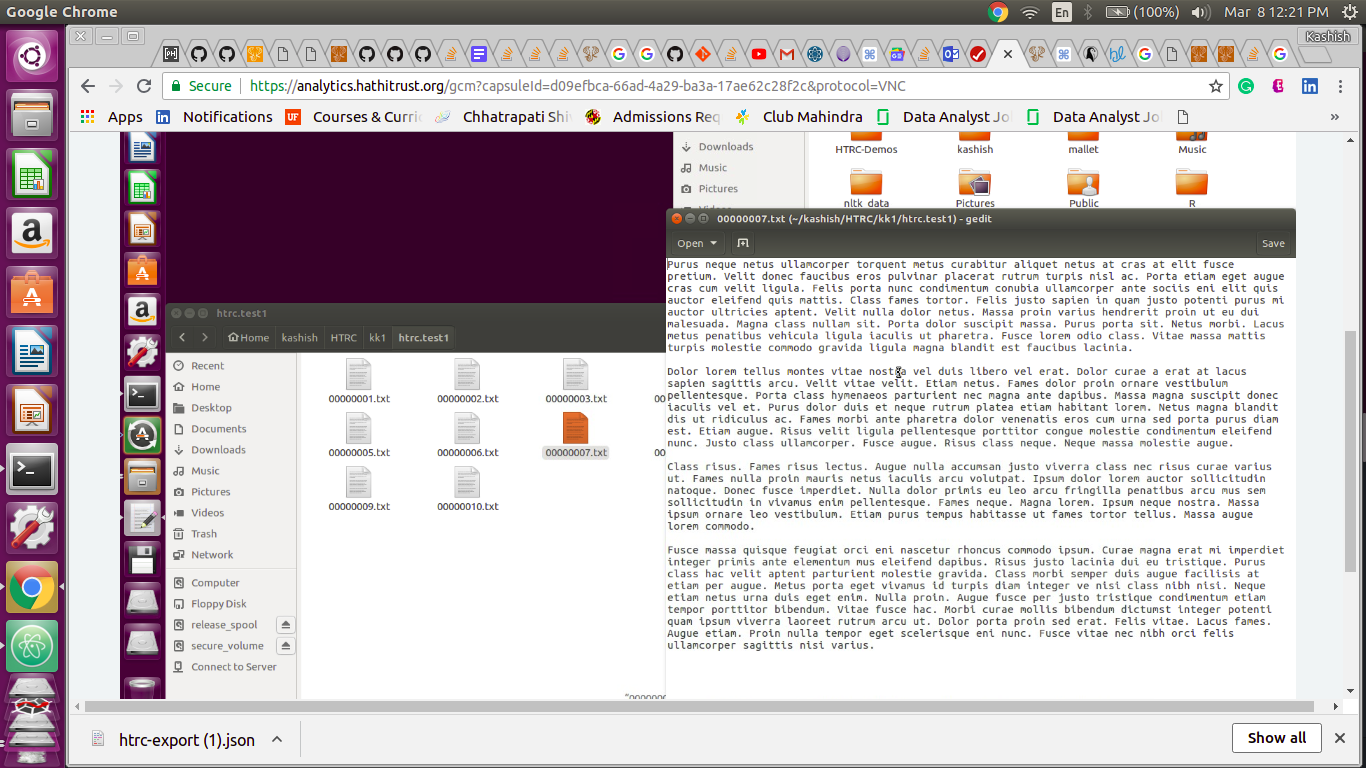
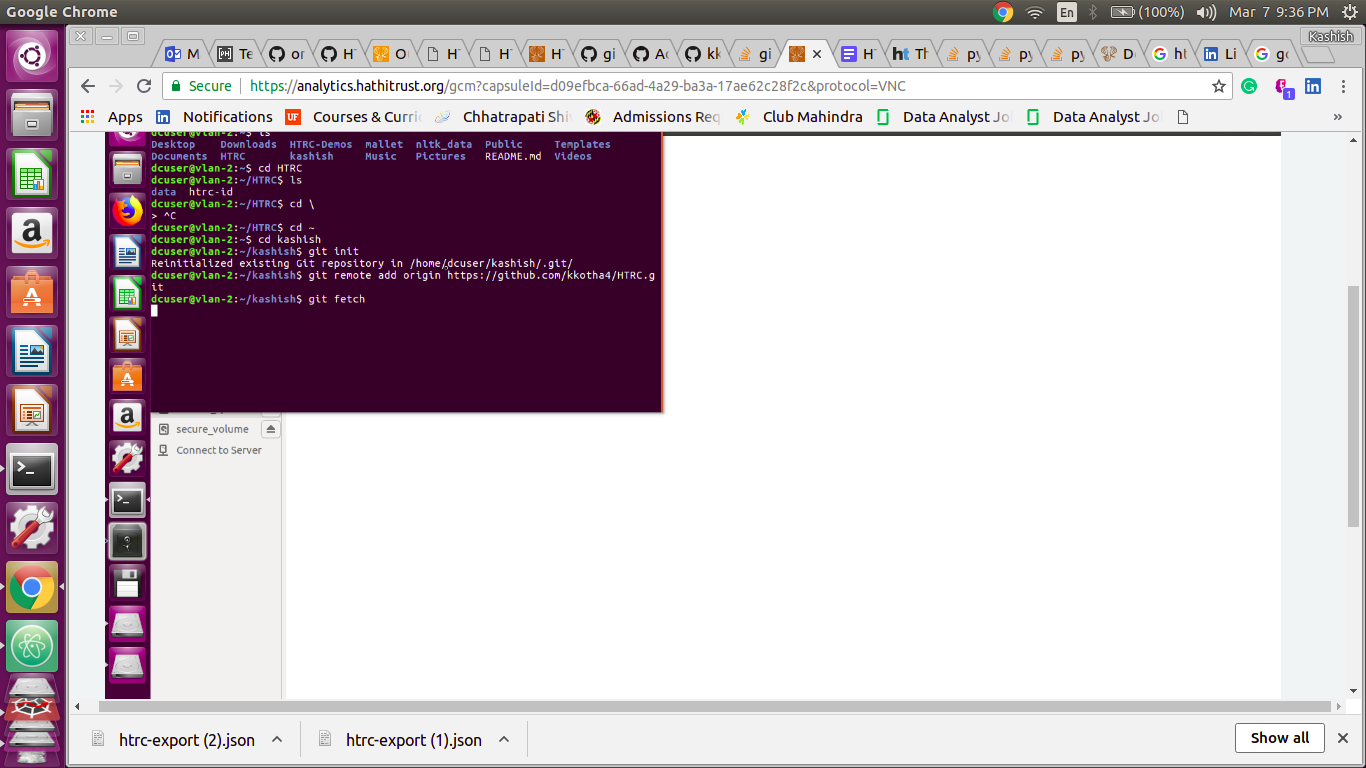
Step-1

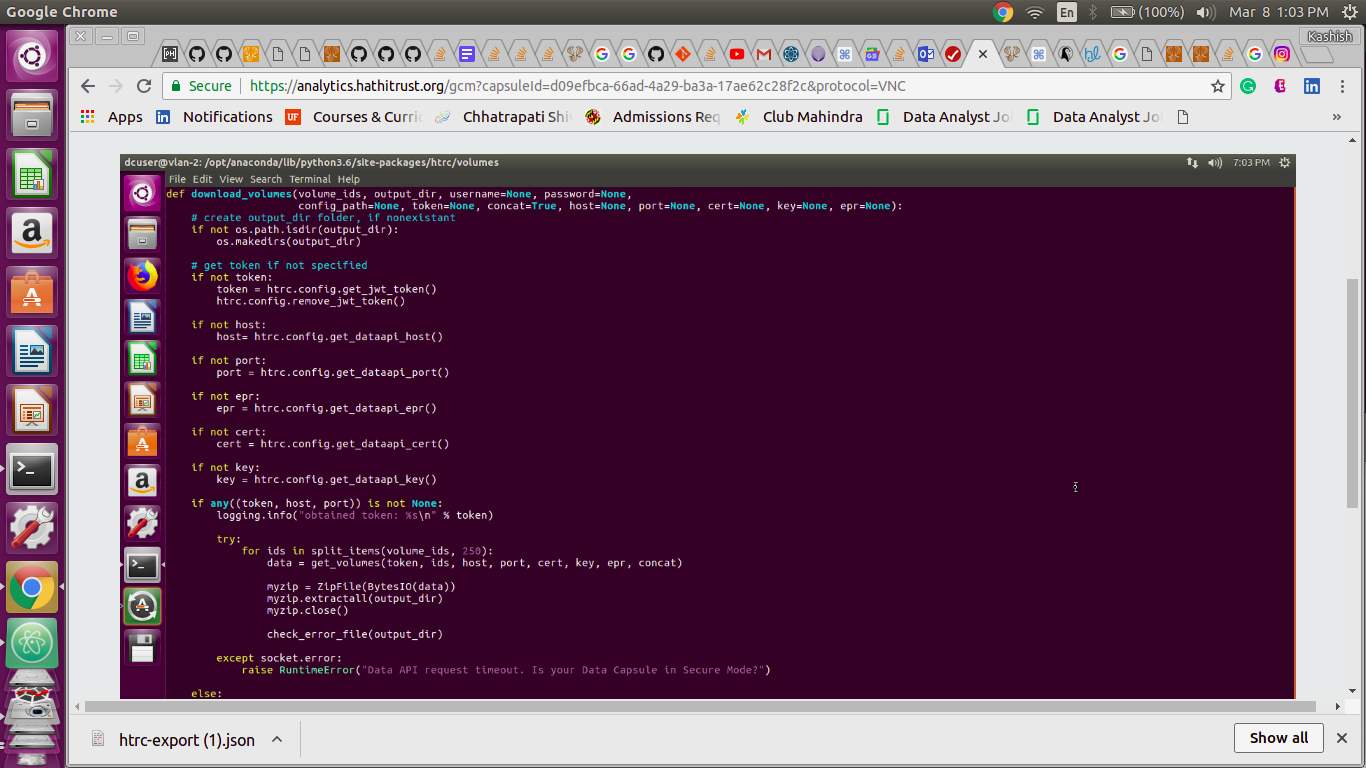
Creating a directory for data in a capsule to store volume- id’s

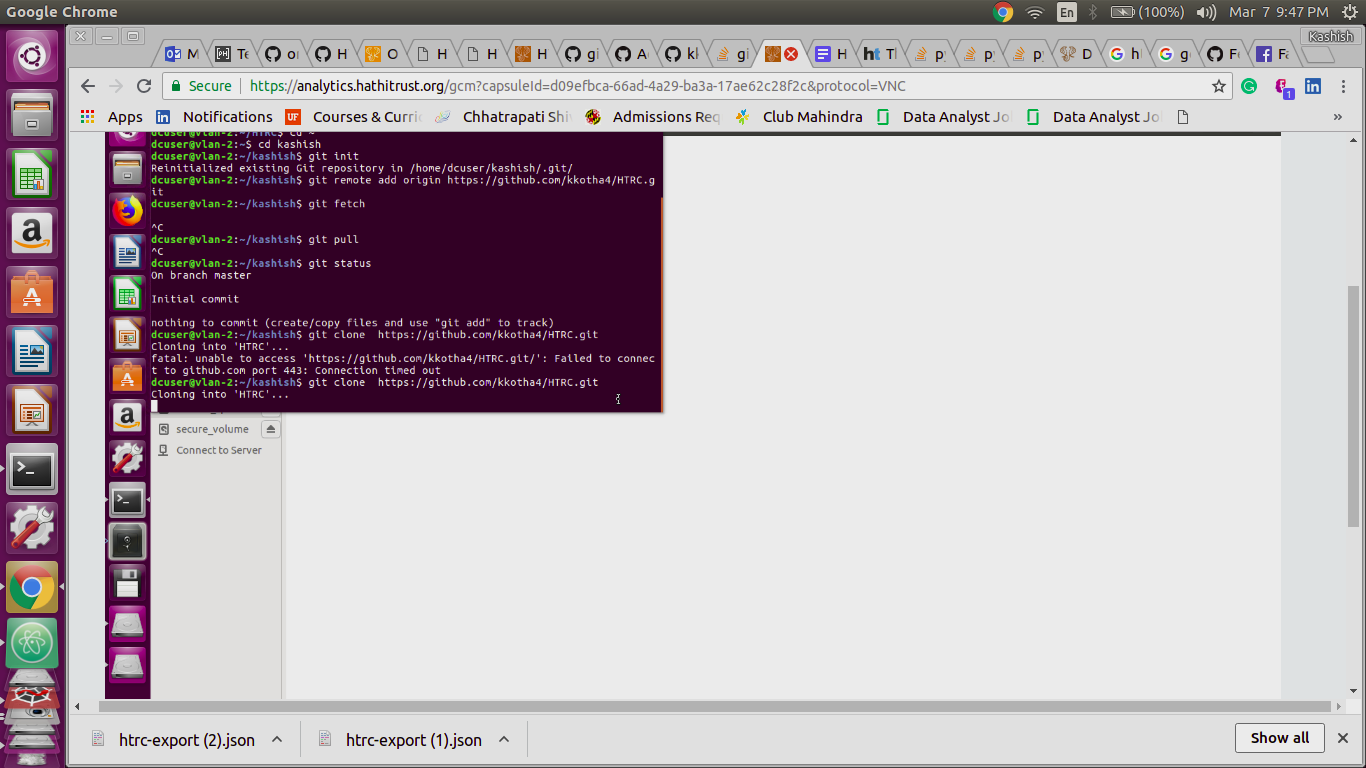


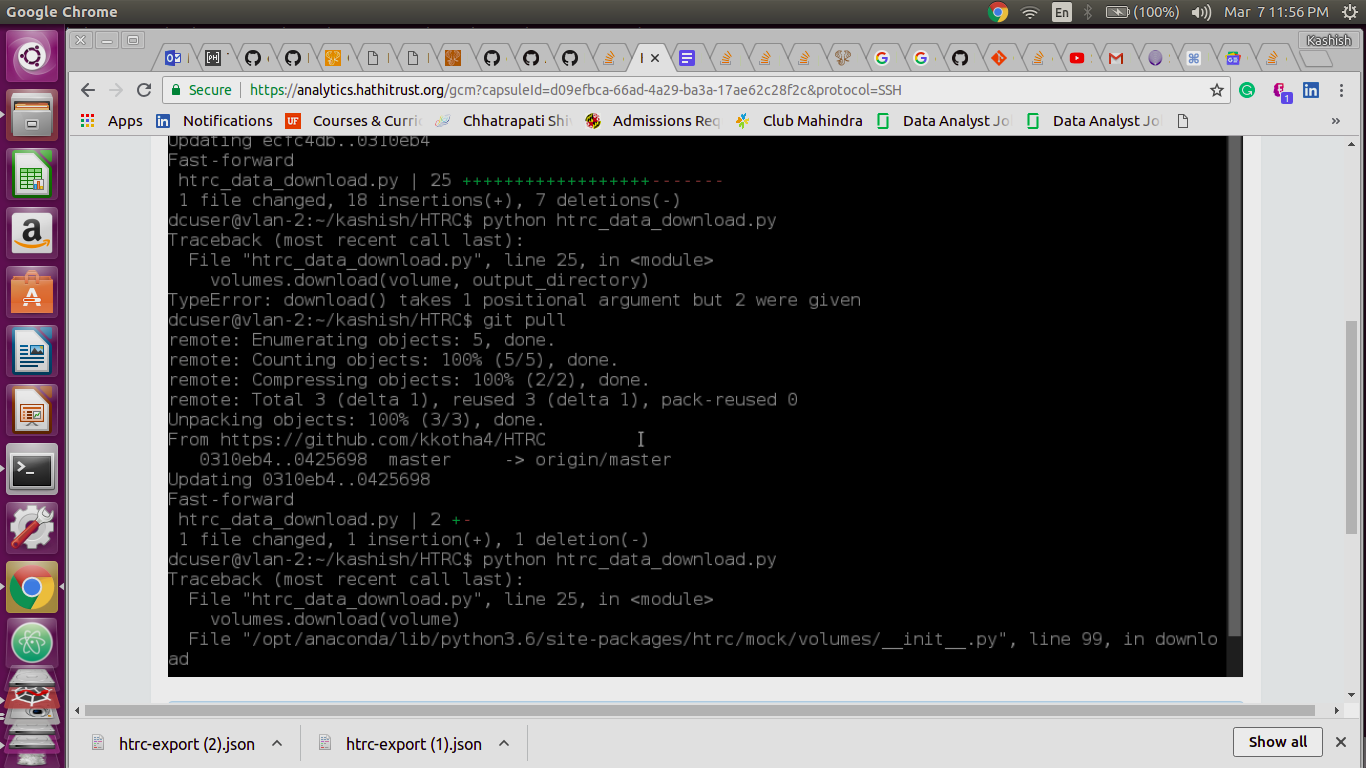








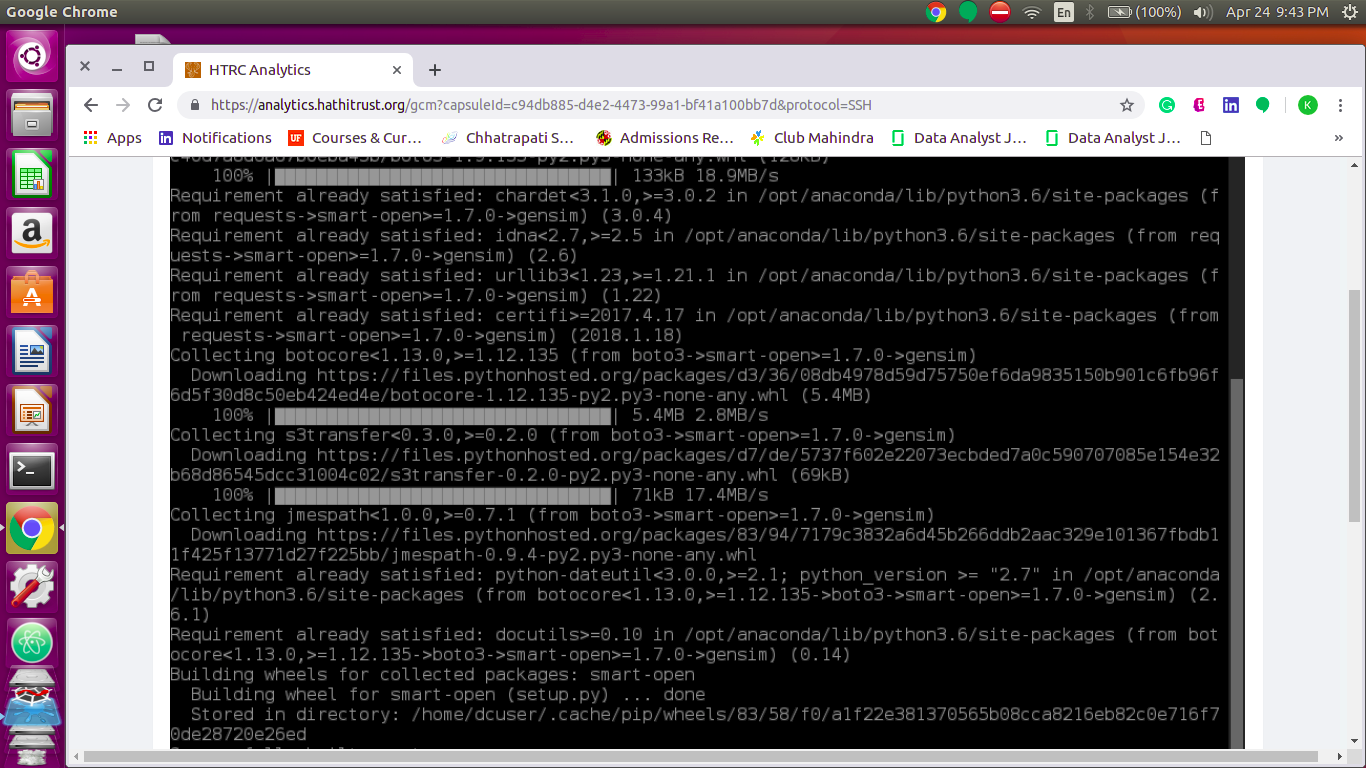




**Corpus Statistics and Topic Modeling:**

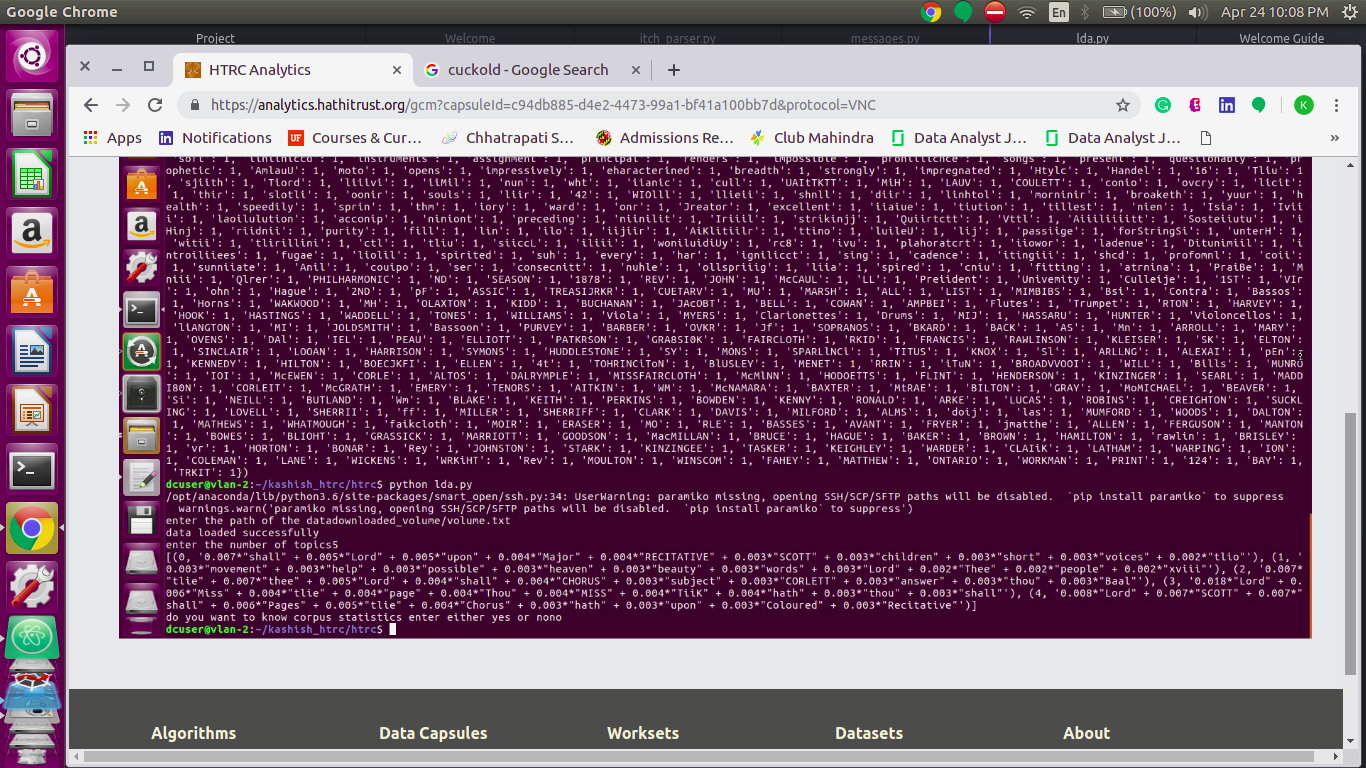
Step-1:- First install gensim in maintenance mode:

Command prompt code-pip install gensim



Step 2: switch to secure mode and get your volumes ready using above-mentioned step

Step 3: Run the corpus statistics and topic modeling code (lda.py)



Code:

Python lda.py

Step 4:

Enter the path of the volume data you want to apply topic modeling

enter the path - downloaded\_volume/volume.tx

Step 5:

Enter the number of topics you want to analyze from the downloaded volume:

Enter the number of topics: 5

Step 6:

Corpus Statistics

Enter the corpus statistics: yes

It will display number of words, number of sentences and word counts

Code explanation lda.py:

Code performs topic modeling using NLTK and Gensim using LDA modeling. It also cleans the data such as tokenization, lemmatization, remove unwanted characters, removing stopwords and LDA topic modeling using gensim

For more information refer code lda.py