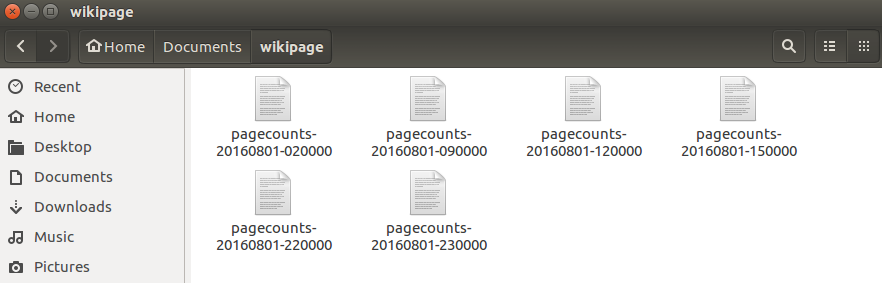
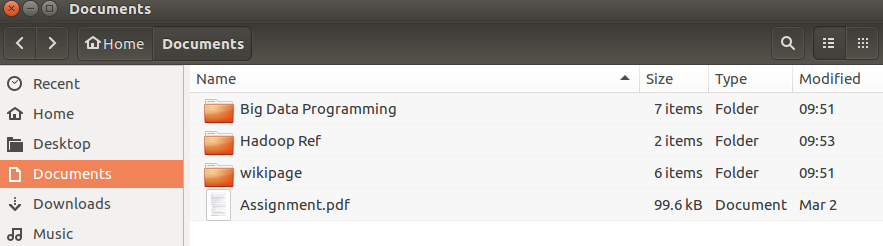
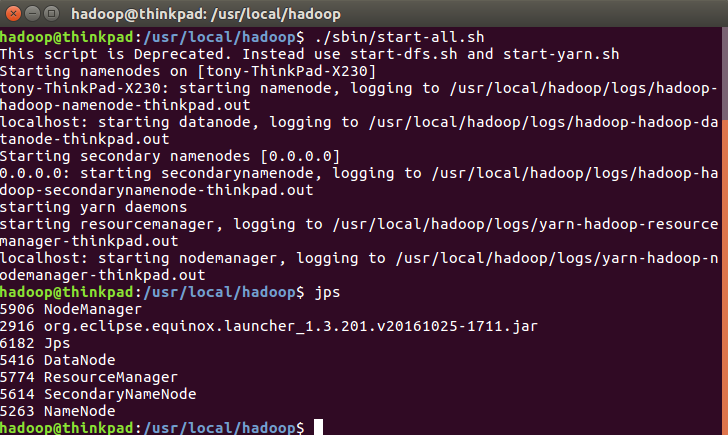
**Loading datasets into HDFS**

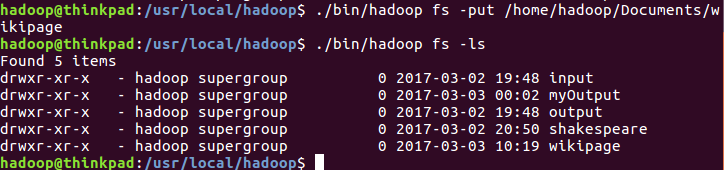
Step1: The xxxx.gz files are unziped and piled together into a single directory. As shown below: wikipage



Step2: Open Linux OS terminal and start hadoop program

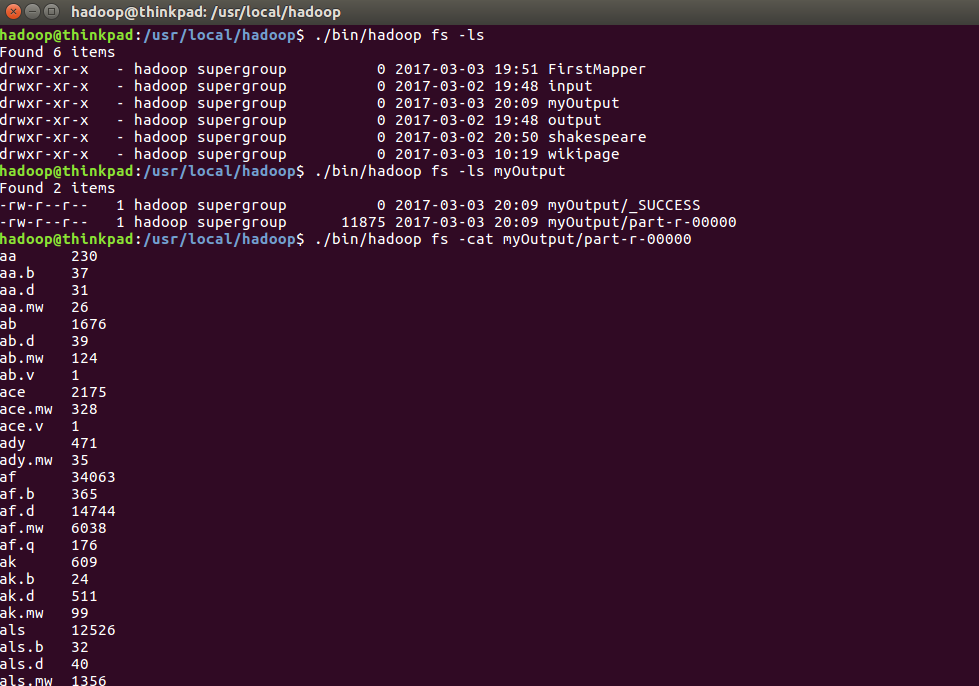


step 3: load wikipage file into HDFS using -put



Task 1 Results

1. Aggregate the number of pageviews based on project name (first column)



2. Sort the results of pageviews based on descending order

