

Advanced Big Data Analytics	
Course Code: UE20CS581	Professor: Dr. K.V Subramaniam
Submitted By: Mehvish Hafiz	SRN: PES1PG20CS021
B. Shruthi Laya	SRN: PES1PG20CS011

Assignment: Creating an Inverted Index for Mahabharata text file.

Github Link: <https://github.com/mehvish24/InvertedIndexMR>

Understanding:

1. Installation of Hadoop & how to start & stop it.
2. Cleaning a text file i.e. removed stop words, punctuation, numbers etc.
3. Splitting the cleaned file into many sections.
4. Creating Directories in Hadoop file system & uploading data onto the file system.
5. Running Inverted Index on the data.
6. Understood how map reduce works.

Output of Inverted Index:

```

user@user-HP-Notebook: ~/Desktop/InvertedIndexMR
xx619=1, xx609=1, xx67=1, xx197=1}
yudhishthir {xx574=5, xx453=3, xx454=2, xx333=1, xx451=2, xx452=2, xx331=1, xx573=1, xx336=3, xx578=2, xx457=7, xx579=2, xx337=4, xx458=4,
xx334=1, xx213=5, xx455=1, xx577=1, xx214=1, xx335=4, xx456=4, xx219=2, xx338=4, xx339=5, xx570=6, xx571=1, xx222=7, xx343=1, xx585=1, xx344=
1, xx220=1, xx583=6, xx584=2, xx342=1, xx94=3, xx221=9, xx468=2, xx469=7, xx345=3, xx466=2, xx467=6, xx346=1, xx588=1, xx349=7, xx581=2, xx582
=7, xx340=2, xx580=3, xx310=5, xx431=2, xx553=3, xx432=3, xx550=4, xx551=1, xx430=1, xx314=2, xx557=7, xx315=5, xx554=1, xx312=5, xx433=6, xx5
5=3, xx434=3, xx318=2, xx439=1, xx319=4, xx558=1, xx316=12, xx438=2, xx317=2, xx201=1, xx440=2, xx567=3, xx325=3, xx447=2, xx568=1, xx323=2,
xx324=9, xx566=1, xx445=1, xx208=6, xx209=2, xx327=1, xx569=2, xx448=2, xx328=1, xx207=4, xx449=1, xx560=2, xx413=1, xx412=3, xx417=1, xx419=1
, xx420=1, xx300=11, xx303=6, xx304=1, xx546=2, xx543=1, xx301=2, xx423=1, xx544=3, xx302=5, xx428=1, xx549=1, xx308=7, xx429=1, xx547=3, xx30
5=4, xx427=4, xx548=3, xx306=1, xx309=2, xx198=3, xx199=2, xx512=9, xx513=4, xx516=2, xx517=1, xx515=1, xx518=1, xx192=1, xx193=3, xx191=6, xx
196=3, xx194=1, xx195=11, xx520=1, xx402=1, xx403=3, xx400=2, xx521=1, xx522=1, xx401=3, xx406=3, xx407=1, xx528=2, xx525=1, xx404=4, xx405=7,
xx526=1, xx529=1, xx409=1, xx298=25, xx611=2, xx291=9, xx292=11, xx171=3, xx290=1, xx295=1, xx296=1, xx293=2, xx294=18, xx189=1, xx187=1, xx1
88=1, xx622=3, xx623=10, xx502=5, xx620=1, xx500=1, xx626=2, xx505=2, xx624=40, xx503=2, xx625=6, xx00=9, xx01=39, xx184=2, xx398=1, xx156=4,
xx278=12, xx399=1, xx157=1, xx155=1, xx279=8, xx43=1, xx270=2, xx273=3, xx394=4, xx152=3, xx271=1, xx392=3, xx150=1, xx151=3, xx272=1, xx393=1
, xx289=3, xx165=1, xx286=6, xx287=2, xx600=3, xx169=2, xx604=2, xx280=4, xx281=4, xx284=6, xx163=5, xx164=1, xx285=10, xx282=6, xx283=4, xx13
5=1, xx132=1, xx374=1, xx60=2, xx495=6, xx133=7, xx61=2, xx254=2, xx138=1, xx62=3, xx259=2, xx139=1, xx499=1, xx258=8, xx493=10, xx372=1, xx25
1=1, xx491=4, xx250=2, xx492=11, xx266=9, xx267=4, xx146=8, xx143=1, xx264=1, xx265=1, xx144=1, xx149=2, xx268=3, xx147=3, xx148=1, xx54=1, xx
269=3, xx262=3, xx383=1, xx142=3, xx260=1, xx381=1, xx261=3, xx140=2, xx475=2, xx476=1, xx473=2, xx474=3, xx479=1, xx237=1, xx238=7, xx114=2,
xx235=3, xx236=7, xx478=3, xx239=5, xx350=3, xx471=3, xx351=1, xx470=5, xx486=1, xx365=6, xx123=1, xx487=1, xx245=1, xx363=2, xx485=1, xx243=2
, xx122=1, xx364=4, xx127=2, xx249=2, xx128=5, xx246=3, xx488=1, xx125=1, xx489=1, xx247=18, xx66=1, xx240=5, xx360=5}
yudhishthir {xx266=5, xx242=1, xx264=3, xx259=2, xx268=2, xx433=1, xx258=1, xx236=1, xx412=1, xx247=2, xx269=4, xx544=1, xx517=1, xx308=1,
xx547=1, xx263=1}
yudhishthir {xx259=1, xx01=1}
yug {xx354=1, xx399=1, xx454=1, xx275=1, xx396=1, xx353=1, xx474=1, xx512=1, xx62=1, xx139=1, xx20=1, xx335=1, xx516=1, xx56=1, xx18=1, xx
443=1, xx501=2, xx325=2, xx502=28, xx31=2, xx53=1, xx445=1, xx225=1, xx500=7, xx624=1, xx22=1, xx66=1, xx00=2, xx24=1, xx25=1, xx461=13, xx340
=1}
yuga {xx245=1, xx583=1, xx62=1, xx501=1, xx512=1, xx237=1, xx500=2, xx00=1, xx01=2, xx316=1, xx503=1, xx461=8}
yugandhar {xx442=1}
yugap {xx122=1}
yup {xx447=1}
yupa {xx447=1}
yuvanash {xx439=4, xx00=1, xx514=2}
yuvuts {xx114=1}
yuyudhan {xx364=1, xx238=1, xx326=1, xx00=1}
yuyuts {xx62=2, xx116=1, xx128=1, xx139=1, xx114=1, xx115=1, xx66=3, xx186=1}
yuyutsh {xx342=1}
zenan {xx239=1}
zodla {xx02=1}
user@user-HP-Notebook:~/Desktop/InvertedIndexMR$ hadoop dfs -cat /InvertedIndexMR/Output/*

```