

Advanced Big Data Analytics

Course Code: UE20CS581

Professor: Dr. K.V Subramaniam

Submitted By:

Mounika N (PES1PG20CS023) Shilpa H (PES1PG20CS052)

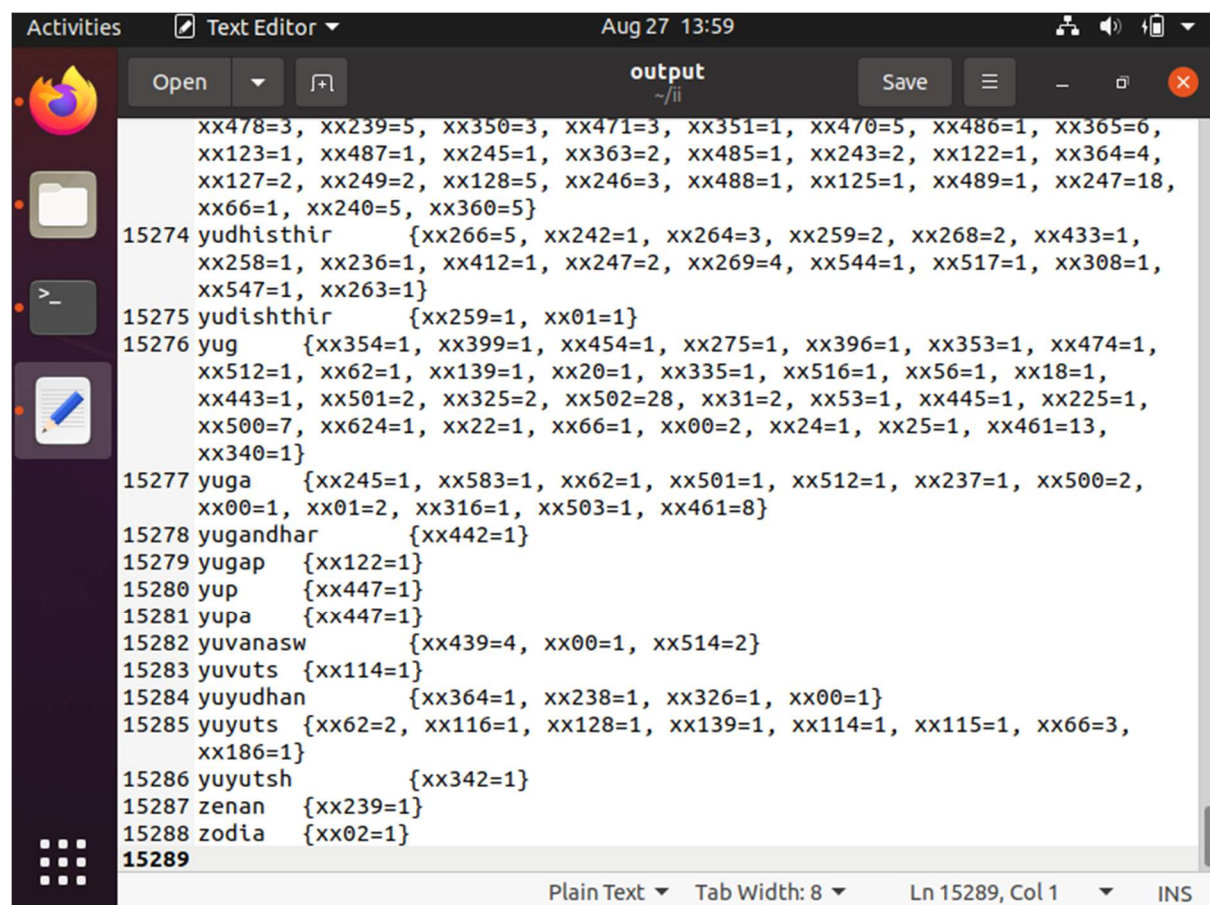
Assignment: Creating an Inverted Index for Mahabharata text file.

Github Link: github.com/shilpa-hegade/InvertedIndex

Learnings:

1. Installation of Hadoop.
2. Preprocessing a large text file.
3. Working of Hadoop file system.
4. Running Inverted Index on the data.

Output of Inverted Index:



```
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}
15274 yudhisthir {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}
15275 yudishthir {xx259=1, xx01=1}
15276 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,
xx443=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}
15277 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}
15278 yugandhar {xx442=1}
15279 yugap {xx122=1}
15280 yup {xx447=1}
15281 yupa {xx447=1}
15282 yuvanaw {xx439=4, xx00=1, xx514=2}
15283 yuvuts {xx114=1}
15284 yuyudhan {xx364=1, xx238=1, xx326=1, xx00=1}
15285 yuyuts {xx62=2, xx116=1, xx128=1, xx139=1, xx114=1, xx115=1, xx66=3,
xx186=1}
15286 yuyutsh {xx342=1}
15287 zenan {xx239=1}
15288 zodia {xx02=1}
15289
```