hadoop@ubuntu22:~$ start-all.sh

hadoop@ubuntu22:~$ jps

hadoop@ubuntu22:~$ gedit weather\_sample.txt

1950 0 1

1950 22 1

1950 -11 1

1949 111 1

1949 78 1

1949 45 0

1951 9999 2

1951 9999 5

1952 9999 9

1953 25 0

hadoop@ubuntu22:~$ pig -x local

grunt>records = LOAD 'weather\_sample.txt' USING PigStorage('\t') AS (year:chararray, temperature:int, quality:int);

grunt> dump records

(1950,0,1)

(1950,22,1)

(1950,-11,1)

(1949,111,1)

(1949,78,1)

(1949,45,0)

(1951,9999,2)

(1951,9999,5)

(1952,9999,9)

(1953,25,0)

grunt> describe records;

records: {year: chararray,temperature: int,quality: int}

grunt> filtered\_records = FILTER records BY temperature!=9999 AND quality IN (0,1,4,5,9);

grunt> dump filtered\_records

(1950,0,1)

(1950,22,1)

(1950,-11,1)

(1949,111,1)

(1949,78,1)

(1949,45,0)

(1953,25,0)

grunt> grouped\_records=GROUP filtered\_records BY year;

grunt> dump grouped\_records

(1949,{(1949,45,0),(1949,78,1),(1949,111,1)})

(1950,{(1950,-11,1),(1950,22,1),(1950,0,1)})

(1953,{(1953,25,0)})

grunt> max\_temp = FOREACH grouped\_records GENERATE group,MAX(filtered\_records.temperature);

grunt> DUMP max\_temp;

(1949,111)

(1950,22)

(1953,25)