

```
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)