**Analysis of Part 01:**

Below is a brief explanation of the amount of time required to run each dataset using the awk script with a graph and chart combined with some screen shots to explain the process in details a compare the time taken by each dataset.

System Configuration:

**Virtual Box Used: Ubuntu/trusty64.**

**Memory allocated: 2048 MB.**

**CPU Speed: 2699.986 MHz**

**Observation:**

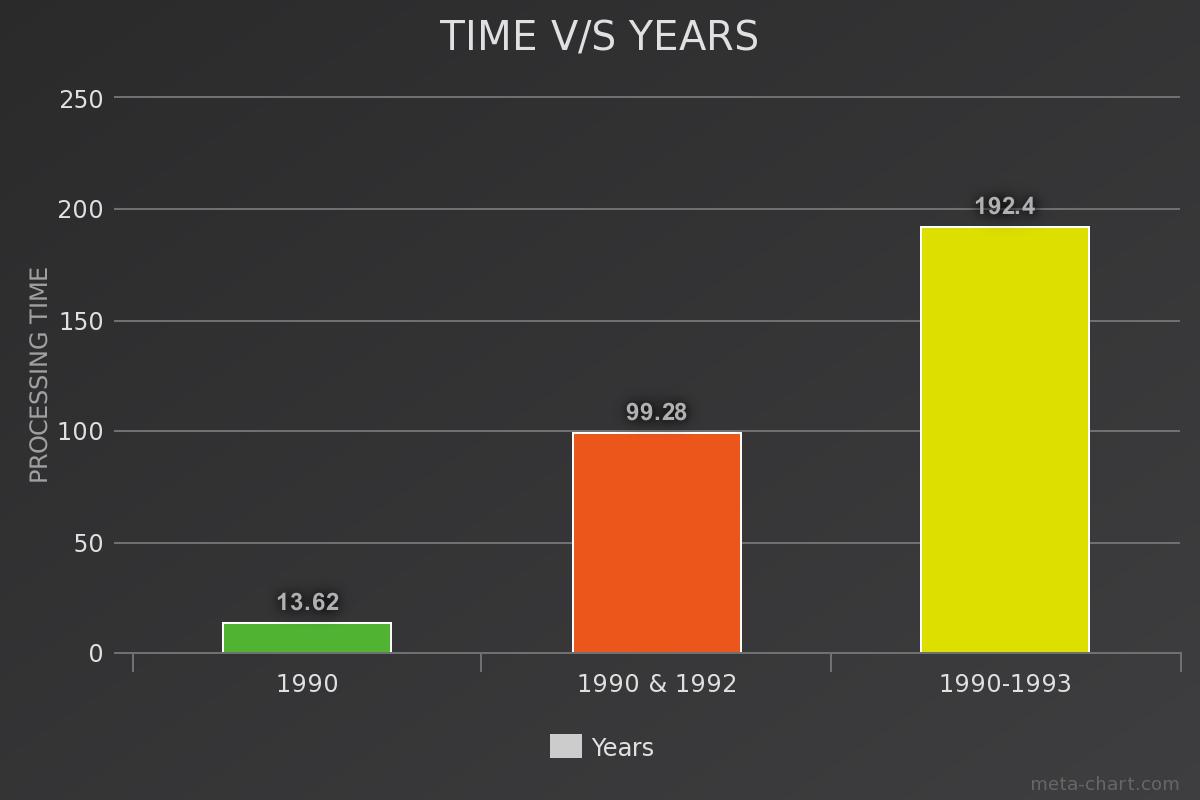
1. Dataset of year 1990 took the least amount of time as compared to the other dataset since the amount of data increased with each dataset. Hence, the time.
2. Processing time of the datasets with memory of 1028 and 2048 was same, concluding that the processing time is not affected by the increase or decrease of memory.
3. The processing time increases while processing huge amount of data.
4. For quicker results and to reduce the processing time, the job can be executed on different processors.
5. The amount Memory free before running the job was 86232 KB, and the Memory free while running the job was 66252 KB (Screen shots attached below is the file)

The below Chart and graph explains about the time taken to run the awk script on each dataset while parsing the huge amount of data in seconds and giving us the max temperature for each year.

**Chart**:

|  |  |  |  |
| --- | --- | --- | --- |
| **Year** | **Time taken to run the job** | **Memory Allocated** | **CPU Speed** |
| 1990 | 13.62 | 2048 MB | 2699.86 |
| 1990 &1992 | 99.28 | 2048 MB | 2699.86 |
| * 1990 -1993 | 192.40 | 2048 MB | 2699.86 |

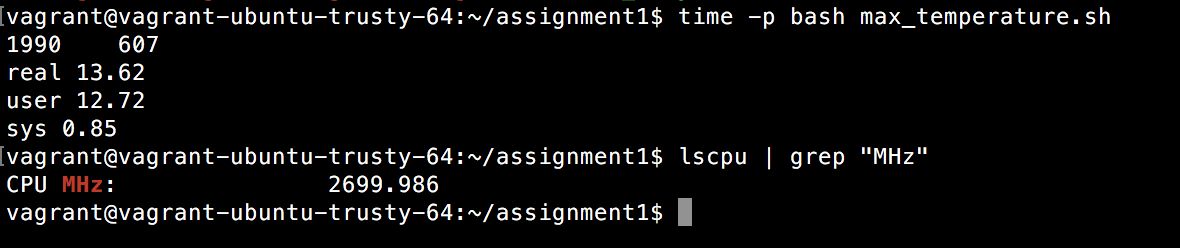
**Graph: Using awk script:**

****

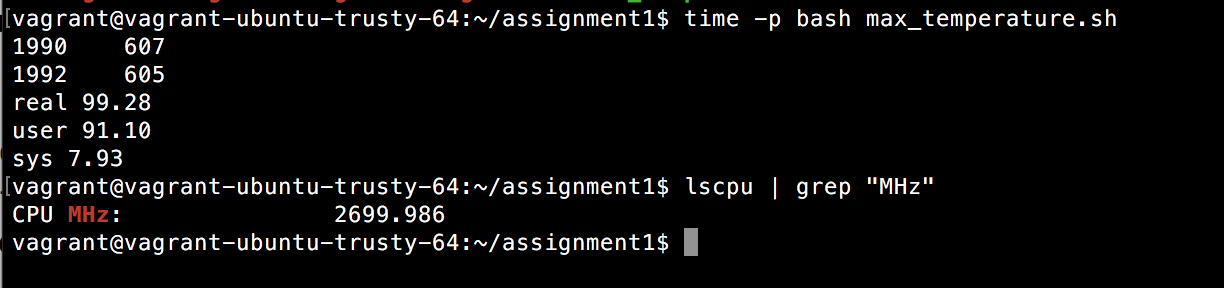
We can see the how quick the awk sciprt parses the entire data and the trend where the time is increased with larger dataset in the above graph while using the awk script.

**Screen Shots:**

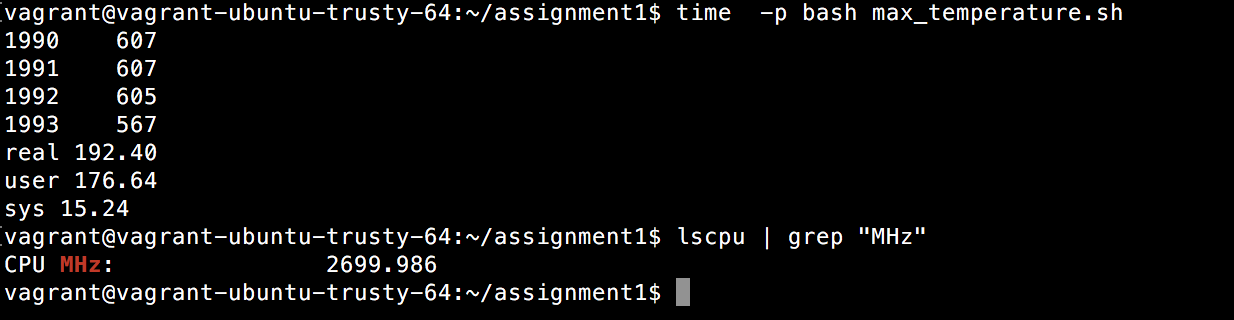
**Result of the dataset 1990 while running the Awk script:**

****

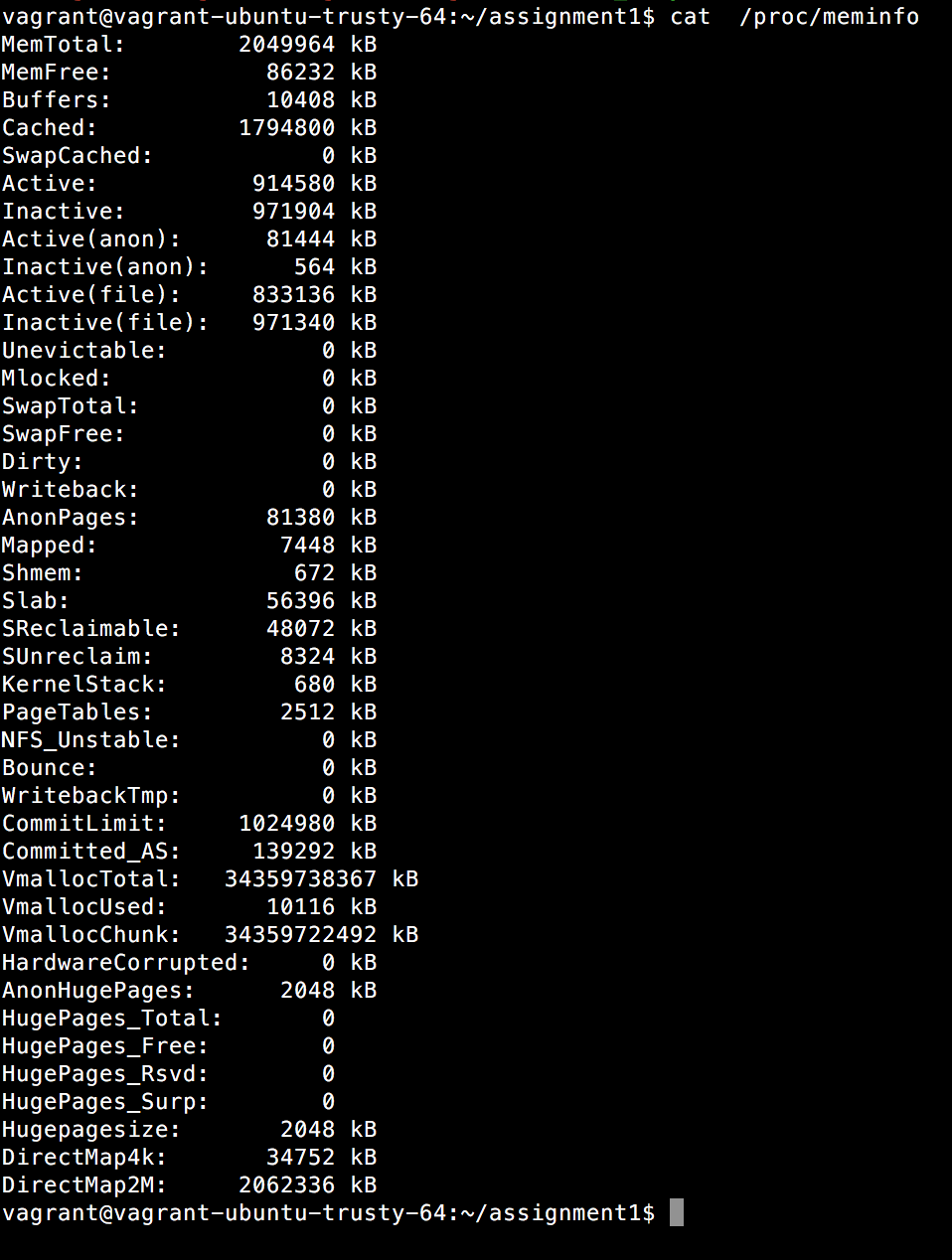
**Result of the dataset 1990 and 1992 while running the Awk script:**

****

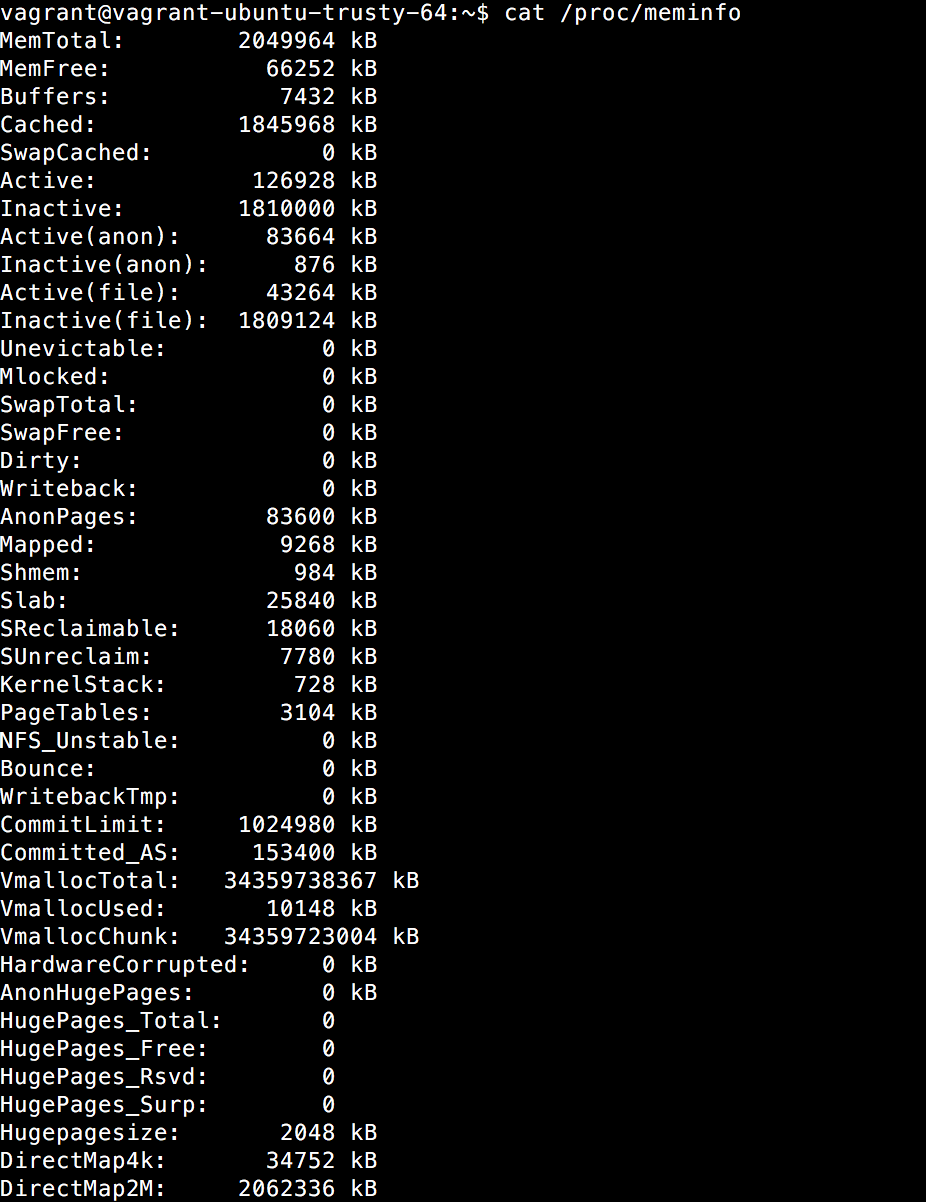
**Result of the dataset 1990-93 while running the Awk script:**

****

**Memory Allocated before running the job:**

****

**Memory Allocated while running the job:**



**Analysis of Part 02:**

In part02, I wrote a java class which will parse each record from the dataset, write it into the CSV file and then load the CSV file into the database executing a SQL query which will return the highest temperature for each year in the database.

Observation:

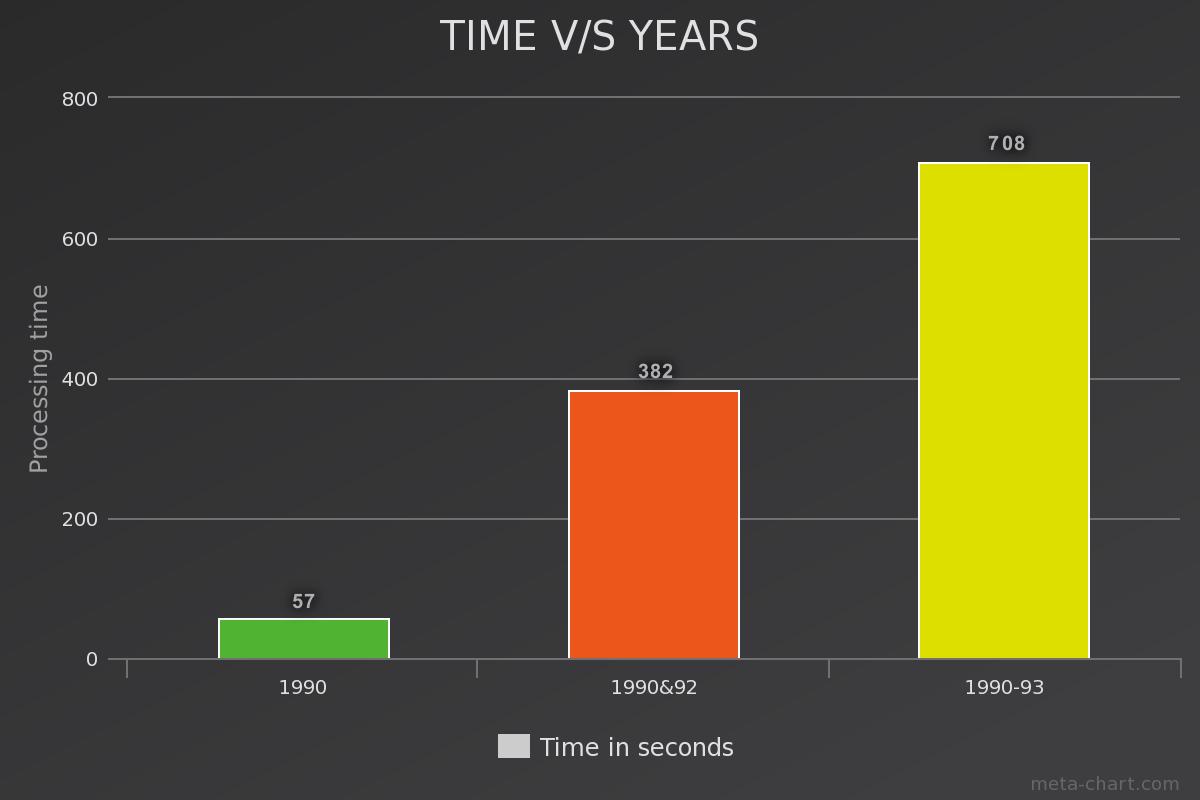
1. Java class was written to parse each dataset and create a csv file. The csv file was then loaded into the DB and then execute a sql query to give me the highest temperature per year. But, the time taken to carry out the entire process took a lot of time as compared to running the awk sciprt on the dataset.
2. But, by improving the code and making it more efficient, I was able to carry out the same operation in less than 50 secs.
3. However, the awk script still seems to be much more efficient to retrieve the data as compared to the Java program.

**Chart:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Year** | **Time taken to run the job in secs** | **Memory Allocated** | **CPU Speed** |
| 1990 | 57 | 2048 MB | 2698.86 |
| 1990 &1992 | 382 | 2048 MB | 2698.86 |
| * 1990 -1993 | 708 | 2048 MB | 2698.86 |

**Graph: Using java:**

We can see the how the java code parses the data and the trend where the time is increased with larger dataset in the above graph while using the java code.

****

**Below is the screen shots of Part02:**

Output in the console showing the highest temperature per year and time taken to execute each dataset.

