Table of Contents

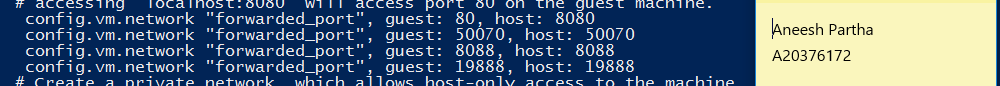
[PART ONE – OBJECTIVE](#_Toc462487186)

[PART TWO – OUTPUT](#_Toc462487187)

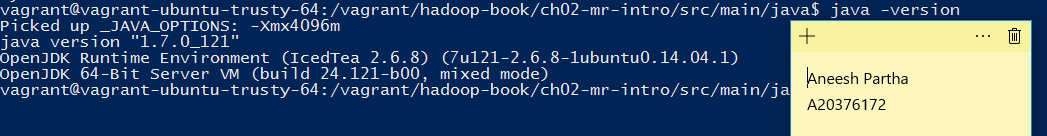
[PART THREE – ASSIGNMENT QUESTIONS](#_Toc462487188)

**Hadoop Installation:**

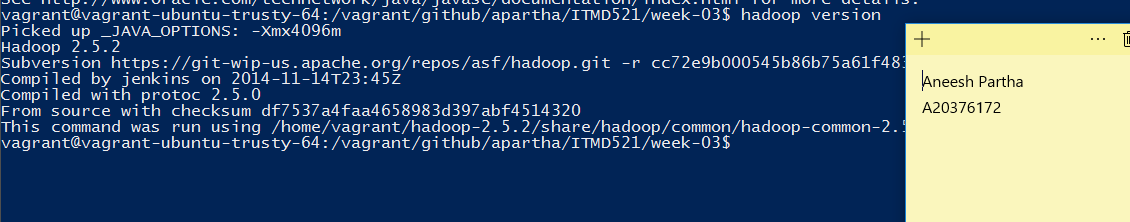
Enabling port:



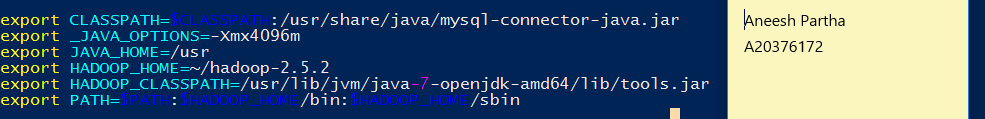
Check for Java availability:



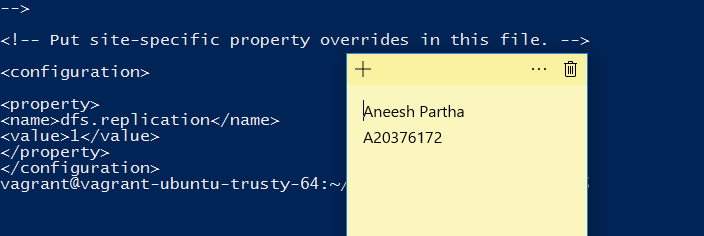
Check after Hadoop is installed in the system:



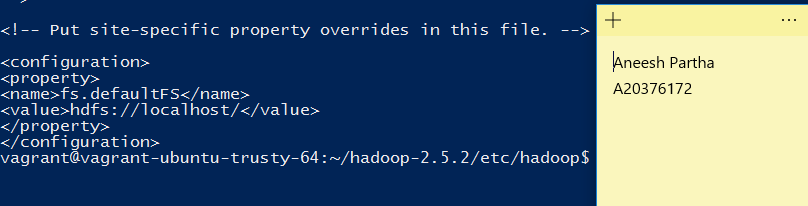
Below paths are set in .bashrc:



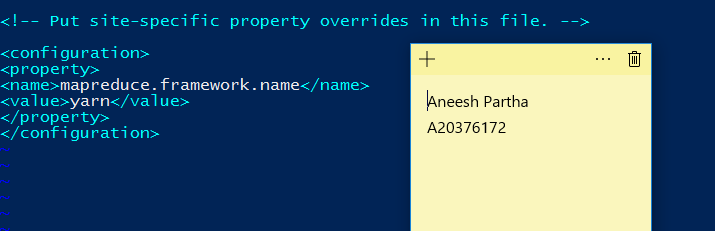
Contents of hdfs-site.xml:



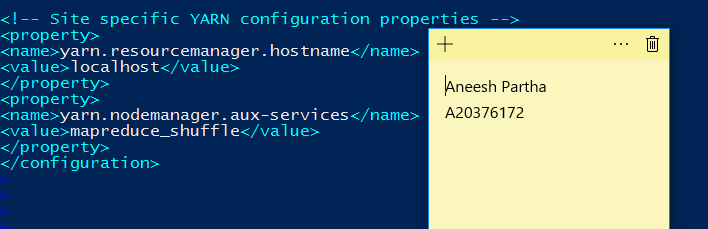
Contents of core-site.xml



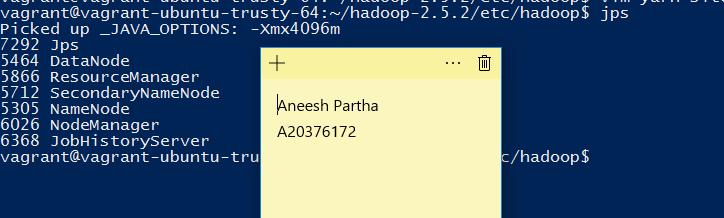
Contents of Mapred-site.xml:



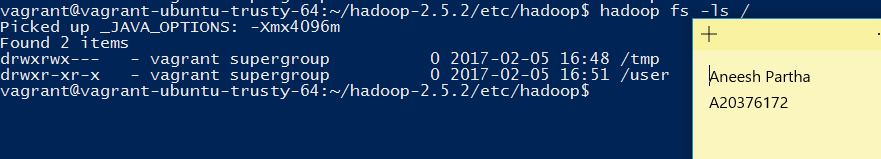
Contents of yarn-site.xml:



Processes running in the system:



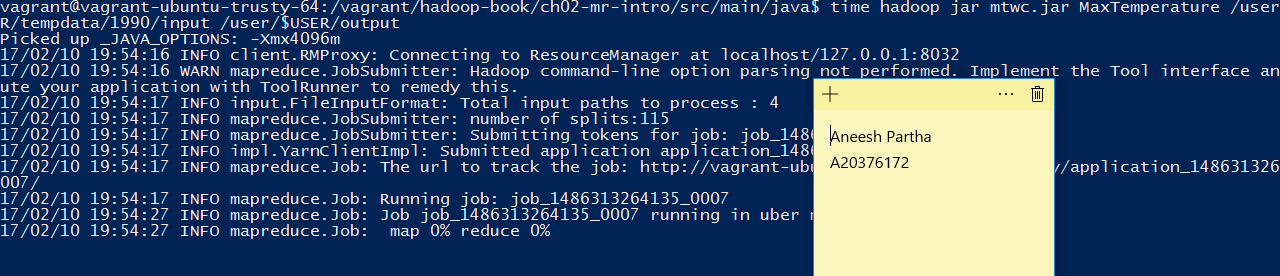
Hadoop file system:

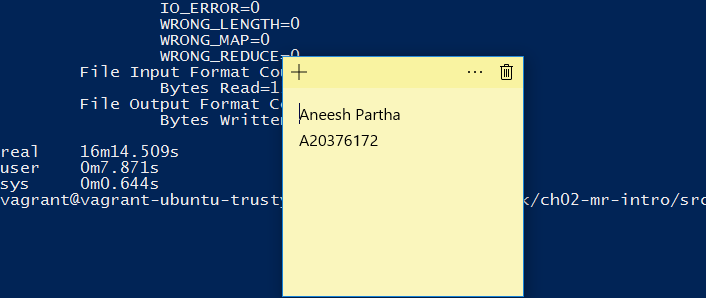


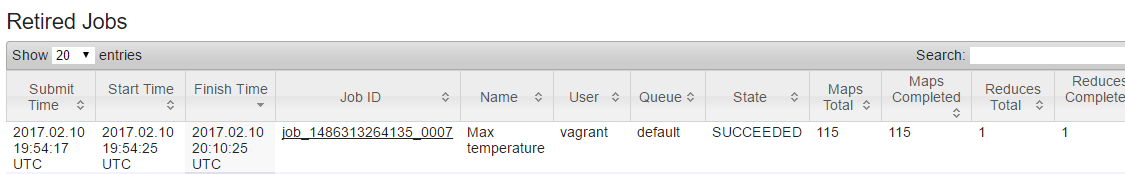
**Analysis of data sets using Hadoop:**

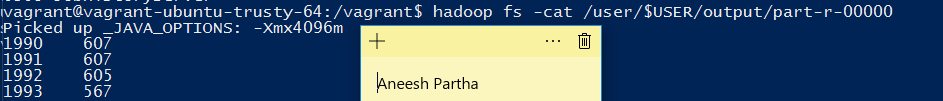
**Without combiner:**

**1990,1991,1992,1993 data sets:**

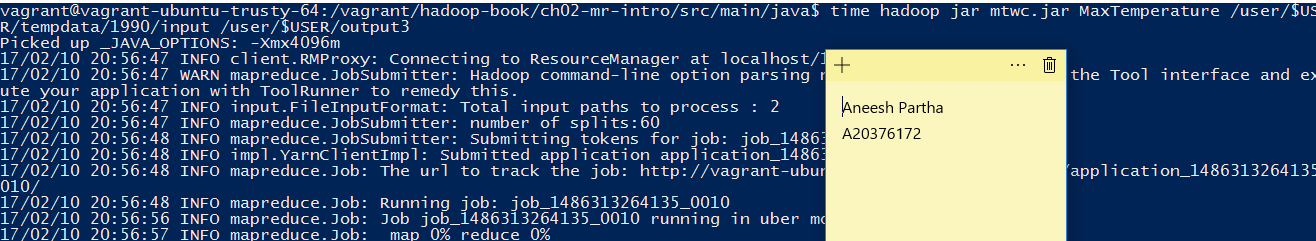


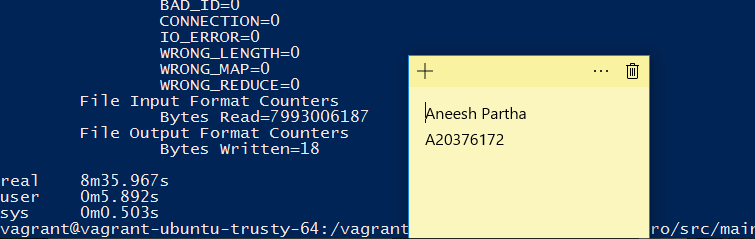


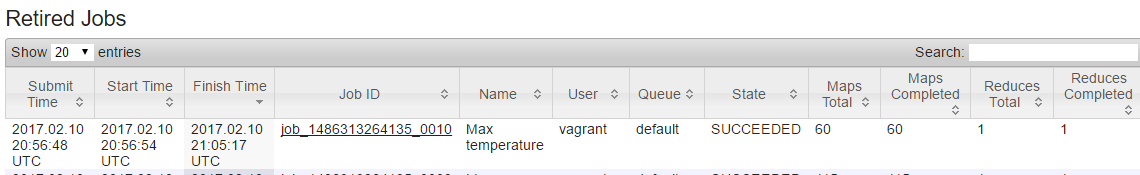


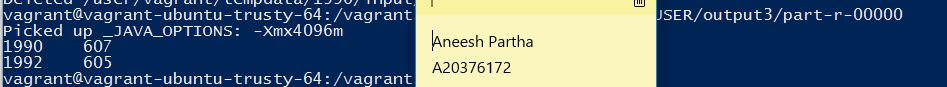


**1990 and 1992 data sets:**

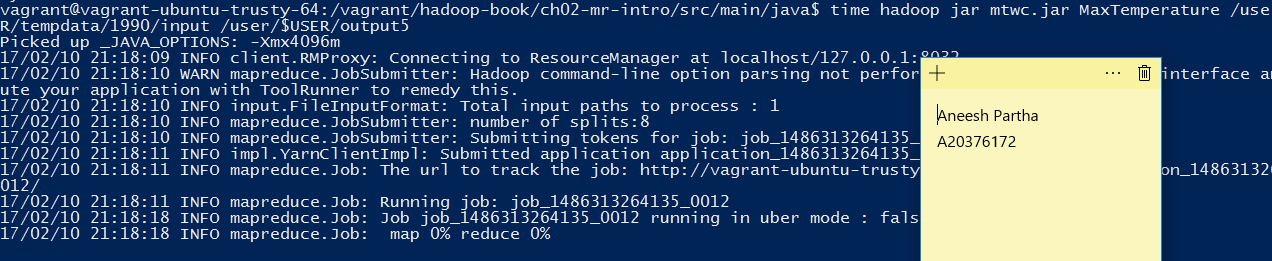


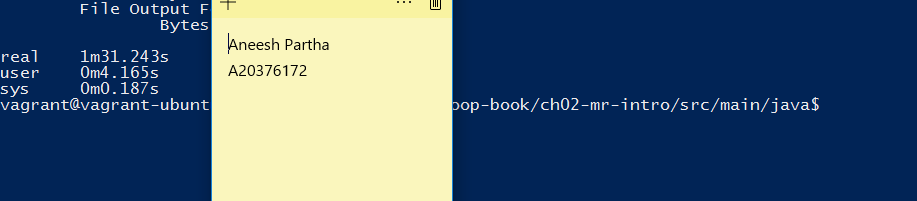


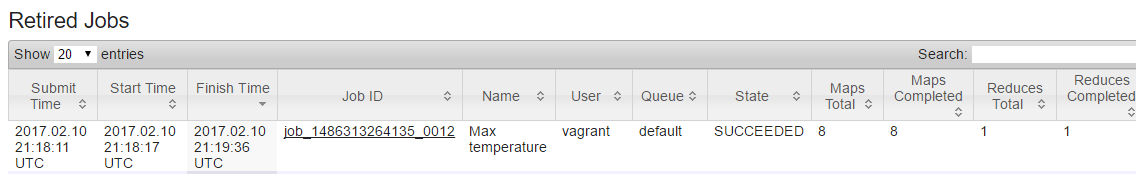


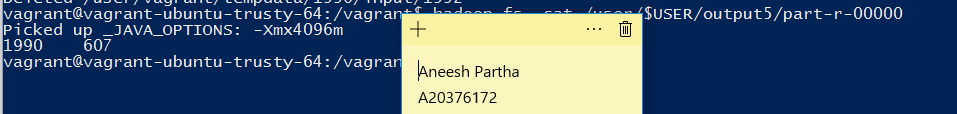


**1990 data set**



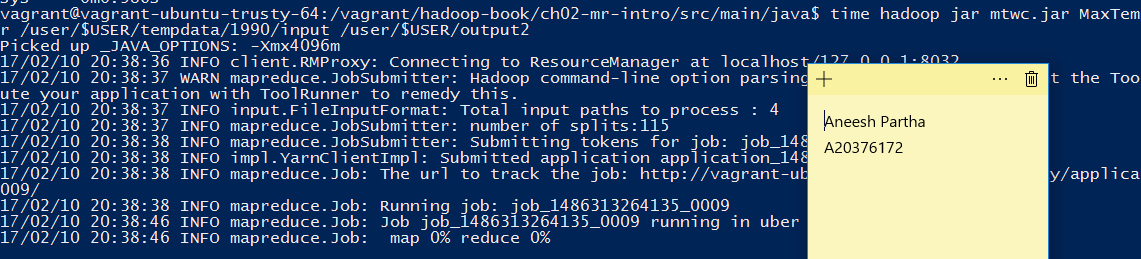


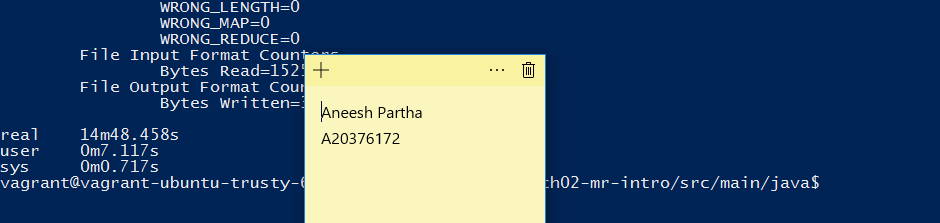


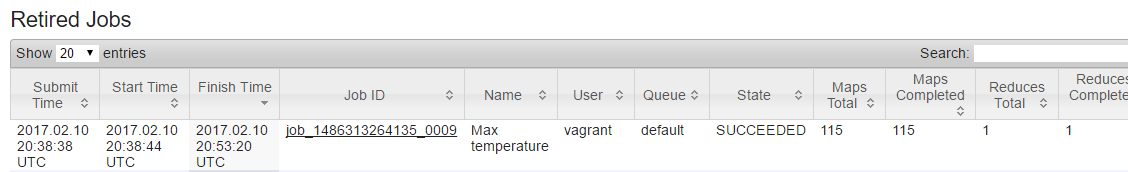


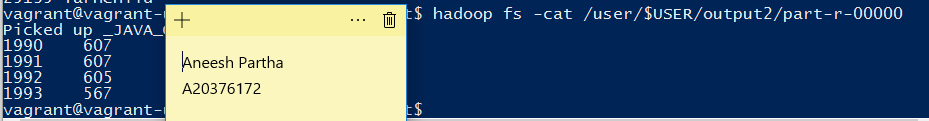
**With combiner:**

**1990,1991,1992 and 1993 data sets**

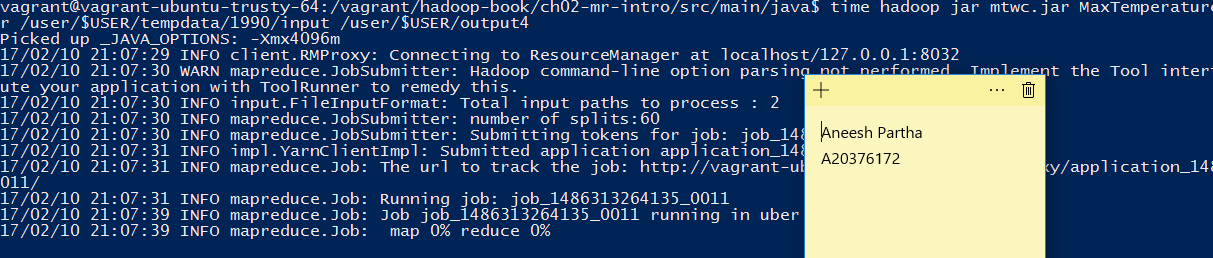


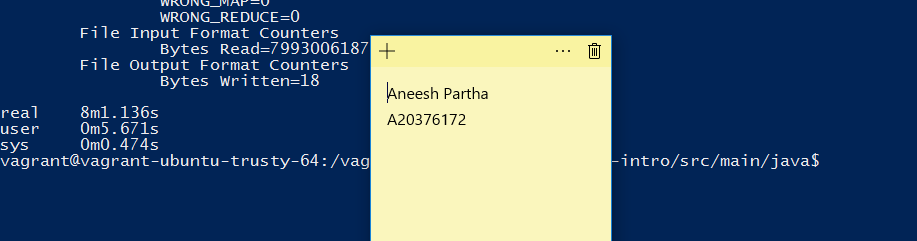


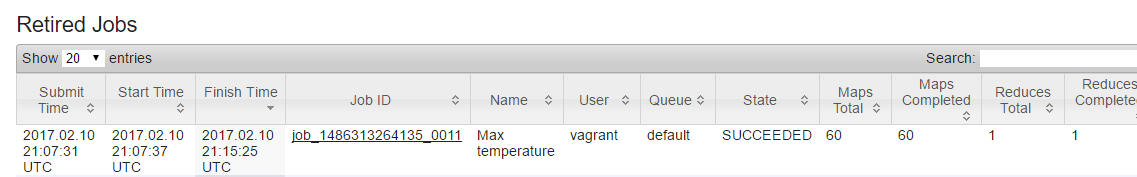


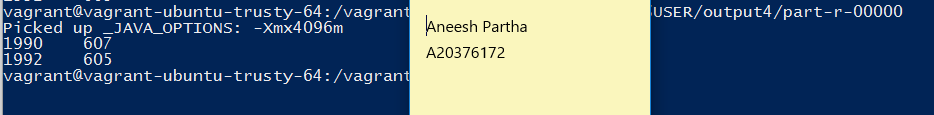


**1990,1992 data sets**



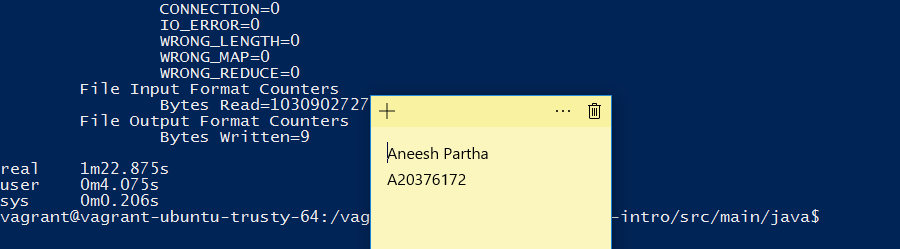


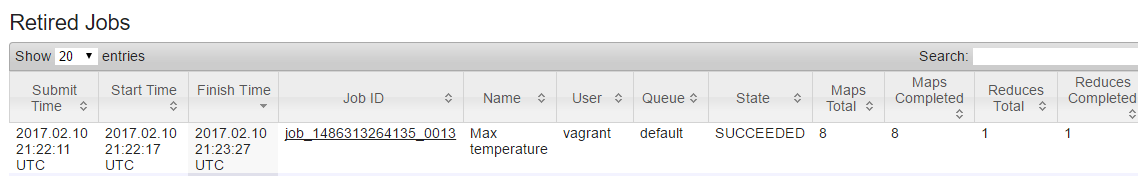


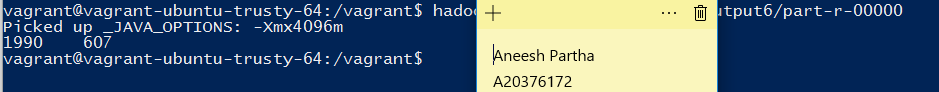


**1990 data set**









**Comparison of time taken:**

|  |  |  |
| --- | --- | --- |
| **Dataset** | **Execution in Hadoop** | **Execution in hadoop using combiner** |
| 1990 | 1.31 | 1.22 |
| 1990 and 1992 | 8.35 | 8.1 |
| 1990,1991,1992 and 1993 | 16.14 | 14.48 |