

|  |  |  |  |
| --- | --- | --- | --- |
| VM | Time(s) | Mins | Secs |
| 2 | 2089 | 34 | 49 |
| 5 | 882 | 14 | 42 |
| 8 | 706 | 11 | 46 |
| 11 | 656 | 10 | 56 |
| 14 | 691 | 11 | 31 |
| 17 | 680 | 11 | 20 |
| 20 | 708 | 11 | 48 |

**Increasing the Number of EC2 instances**

More VM’s will decrease the execution workflow time.

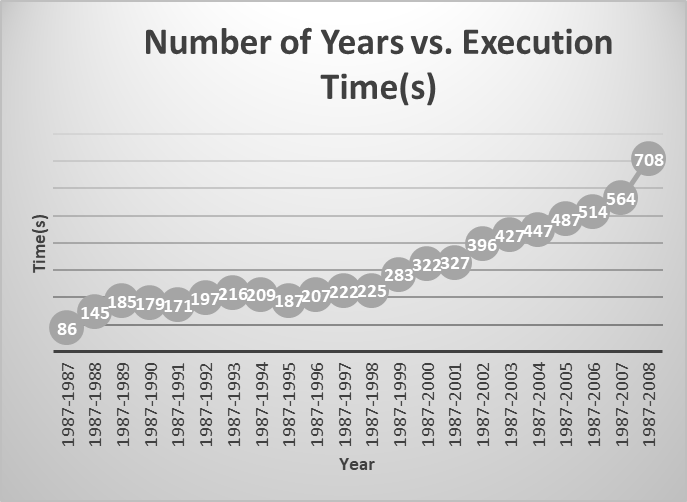
Execution time of each Map-Reduce job will be shorter given more instances.

Flats out after 17 and starts to take more time to execute.

This is because more VM’s == more information interaction time between data nodes.

Information interaction time in Hadoop cluster increases when the number of instances increase.

Scalability: <https://www.geeksforgeeks.org/hadoop-cluster-properties-and-its-types/>



|  |  |  |
| --- | --- | --- |
| Year | Time(s) |  |
| 1987-1987 | 86 | 1 minutes 26 seconds |
| 1987-1988 | 145 | 2 minutes 25 seconds |
| 1987-1989 | 185 | 3 minutes 05 seconds |
| 1987-1990 | 179 | 2 minutes 59 seconds |
| 1987-1991 | 171 | 2 minutes 51 seconds |
| 1987-1992 | 197 | 3 minutes 17 seconds |
| 1987-1993 | 216 | 3 minutes 36 seconds |
| 1987-1994 | 209 | 3 minutes 29 seconds |
| 1987-1995 | 187 | 3 minutes 07 seconds |
| 1987-1996 | 207 | 3 minutes 27 seconds |
| 1987-1997 | 222 | 3 minutes 42 seconds |
| 1987-1998 | 225 | 3 minutes 45 seconds |
| 1987-1999 | 283 | 4 minutes 43 seconds |
| 1987-2000 | 322 | 5 minutes 22 seconds |
| 1987-2001 | 327 | 5 minutes 27 seconds |
| 1987-2002 | 396 | 6 minutes 36 seconds |
| 1987-2003 | 427 | 7 minutes 07 seconds |
| 1987-2004 | 447 | 7 minutes 27 seconds |
| 1987-2005 | 487 | 8 minutes 07 seconds |
| 1987-2006 | 514 | 8 minutes 34 seconds |
| 1987-2007 | 564 | 9 minutes 24 seconds |
| 1987-2008 | 708 | 11 minutes 48 seconds |

**Increasing data size**

Increasing data size will also increase execution time.

From 1998 to 2008, flight data captured increases. This explains the increase in execution time.

More people are also travelling by plane