Google Cluster Trace Analysis with Apache Spark

In this work, I analysed Google cluster-usage traces using Apache Spark and I created the document using R Studio.

You can access the subject of this work from this link. You can find the detailed documantatin of traces from this link.

This version of the code does not contains the Spark codes which are used to extract information from the traces also this is not the final version. There is no guarante related to correctness of the results.

Machine Distibution

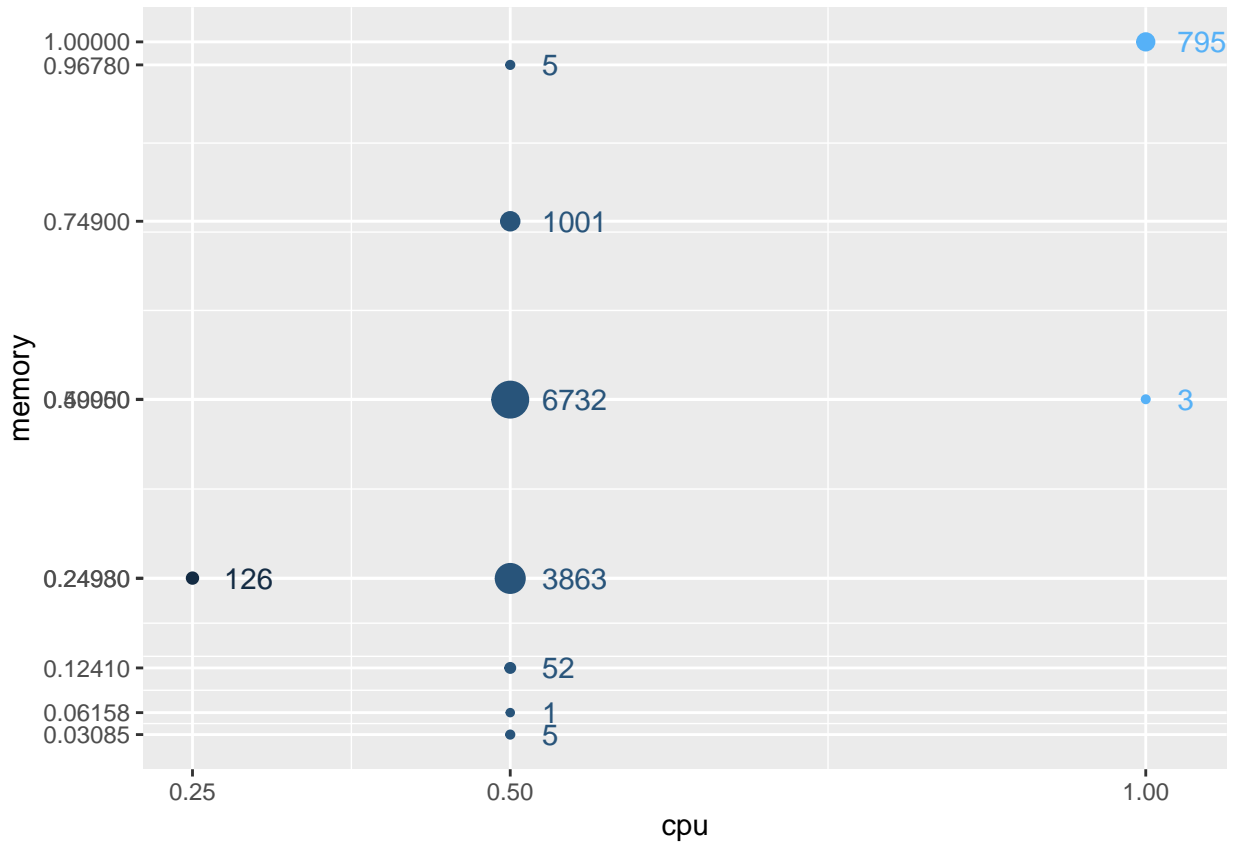## Machine Distibution According to CPU and Memory Capacity

Table 1: Tabular Data

cpu	memory	capacity	number_of_machines	machine_percentage
1.00	1.00000	1.000000	795	6.32
1.00	0.50000	0.750000	3	0.02
0.50	0.96780	0.733900	5	0.04
0.50	0.74900	0.624500	1001	7.96
0.50	0.49950	0.499750	6732	53.50
0.50	0.24930	0.374650	3863	30.70
0.50	0.12410	0.312050	52	0.41
0.50	0.06158	0.280790	1	0.01
0.50	0.03085	0.265425	5	0.04
0.25	0.24980	0.249900	126	1.00

Machine Distibution According to Platform

platform_id	number_of_machines
HofLGzk1Or/8Ildj2+Lqv0UGGvY82NLoni8+J/Yy0RU=	11659
GtXakjpd0CD41brK7k/27s3Eby3RpJKy7taB9S8UQRA=	798
70ZOvysYGtB6j9MUHMPzA2Iy7GRzWeJTdX0YCLRKGVg=	126