In [2]:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import warnings
```

Timing Plot

In [174]:

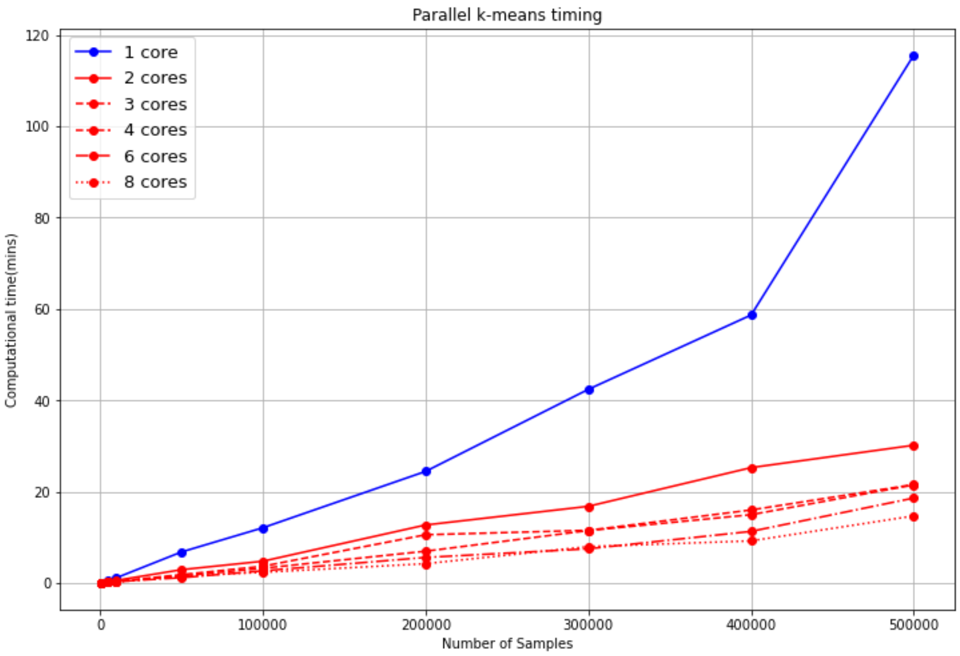
```
df = pd.read_csv('Timing data.csv')
df
```

Out[174]:

	No.of Samples	Serial	2 Cores	3 Cores	4 Cores	6 Cores	8 Cores
0	500	0.032804	0.169684	0.124049	0.108153	0.108821	0.092901
1	1000	0.127690	0.169807	0.158088	0.147818	0.177658	0.183621
2	5000	0.678620	0.398550	0.233278	0.304865	0.224854	0.227826
3	10000	1.176799	0.621300	0.433099	0.327972	0.371066	0.251984
4	50000	6.835195	2.960468	1.874553	1.582210	1.208425	1.583877
5	100000	12.130214	4.811345	3.708052	3.271361	2.709837	2.388150
6	200000	24.466023	12.742547	10.604429	7.002830	5.640169	4.272472
7	300000	42.419725	16.824702	11.573358	11.523497	7.530621	8.012520
8	400000	58.680946	25.289380	16.023895	14.995534	11.344406	9.266653
9	500000	115.566191	30.192446	21.599561	21.516103	18.645391	14.726964

In [181]:

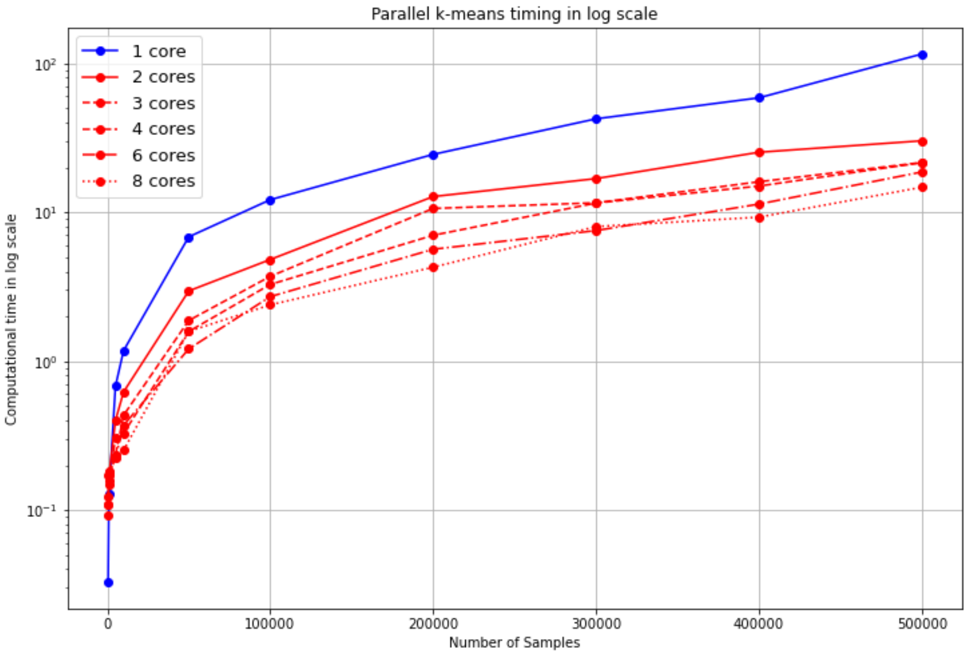
```
plt.figure(figsize=(12,8))
plt.plot(df['No.of Samples'],df['Serial'],'-b',marker='o', label='1 core')
plt.plot(df['No.of Samples'],df['2 Cores'],'-r',linestyle = 'solid',marker='o', label='2 cores')
plt.plot(df['No.of Samples'],df['3 Cores'],'-r',linestyle = '--',marker='o', label='3 cores')
plt.plot(df['No.of Samples'],df['4 Cores'],'-r',linestyle = 'dashed',marker='o', label='4 cores')
plt.plot(df['No.of Samples'],df['6 Cores'],'-r',linestyle = 'dashdot',marker='o', label='6 cores')
plt.plot(df['No.of Samples'],df['8 Cores'],'-r',linestyle = 'dotted',marker='o', label='8 cores')
plt.xlabel('Number of Samples')
plt.ylabel('Computational time(mins)')
plt.title('Parallel k-means timing')
plt.legend(loc='upper left',prop={'size': 13})
plt.grid()
plt.show()
```



Timing Plot in Log Scale

In [24]:

```
plt.figure(figsize=(12,8))
plt.plot(df['No.of Samples'],df['Serial'],'-b',marker='o', label='1 core')
plt.plot(df['No.of Samples'],df['2 Cores'],'-r',linestyle = 'solid',marker='o', label='2 cores')
plt.plot(df['No.of Samples'],df['3 Cores'],'-r',linestyle = '--',marker='o', label='3 cores')
plt.plot(df['No.of Samples'],df['4 Cores'],'-r',linestyle = 'dashed',marker='o', label='4 cores')
plt.plot(df['No.of Samples'],df['6 Cores'],'-r',linestyle = 'dashdot',marker='o', label='6 cores')
plt.plot(df['No.of Samples'],df['8 Cores'],'-r',linestyle = 'dotted',marker='o', label='8 cores')
plt.xlabel('Number of Samples')
plt.ylabel('Computational time in log scale')
plt.title('Parallel k-means timing in log scale')
plt.legend(loc='upper left',prop={'size': 13})
plt.grid()
plt.yscale("log")
plt.show()
```



SpeedUp Curve

In [144]:

```
data = pd.read_csv('Timing Data - SpeedUp.csv')
```

In [145]:

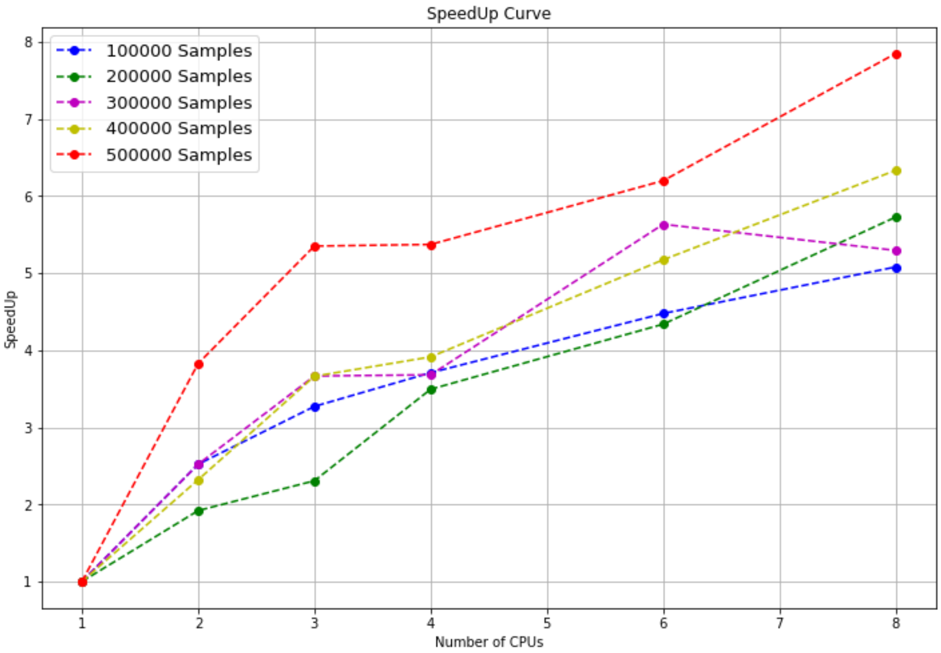
```
data
```

Out[145]:

	No.of Cores	500 Samples	1000 Samples	5000 Samples	10000 Samples	50000 Samples	100000 Samples	200000 Samples	300000 Samples	400000 Samples	500000 Samples
0	1	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000
1	2	0.193322	0.751971	1.702723	1.894093	2.308823	2.521169	1.920026	2.521276	2.320379	3.827652
2	3	0.264441	0.807715	2.909057	2.717163	3.646307	3.271317	2.307151	3.665291	3.662090	5.350395
3	4	0.303307	0.863832	2.225972	3.588105	4.320029	3.708002	3.493734	3.681150	3.913228	5.371149
4	6	0.301446	0.718742	3.018046	3.171403	5.656283	4.476363	4.337817	5.632965	5.172677	6.198110
5	8	0.353102	0.695401	2.978683	4.670144	4.315484	5.079336	5.726433	5.294180	6.332486	7.847251

In [162]:

```
plt.figure(figsize=(12,8))
plt.plot(data['No.of Cores'], data['100000 Samples'], '-b',linestyle = '--', marker='o', label='100000 Samples')
plt.plot(data['No.of Cores'], data['200000 Samples'], '-g',linestyle = '--', marker='o', label='200000 Samples')
plt.plot(data['No.of Cores'], data[' 300000 Samples'], '-m',linestyle = '--', marker='o', label='300000 Samples')
plt.plot(data['No.of Cores'], data['400000 Samples'], '-y',linestyle = '--', marker='o', label='400000 Samples')
plt.plot(data['No.of Cores'], data['500000 Samples'], '-r',linestyle = '--', marker='o', label='500000 Samples')
plt.xlabel('Number of CPUs')
plt.ylabel('SpeedUp')
plt.title('SpeedUp Curve')
plt.legend(loc='upper left',prop={'size': 13})
plt.grid()
plt.show()
```



SizeUp Curve

In [29]:

```
speedup_data = pd.read_csv('Timing Data - SizeUp.csv')
```

In [30]:

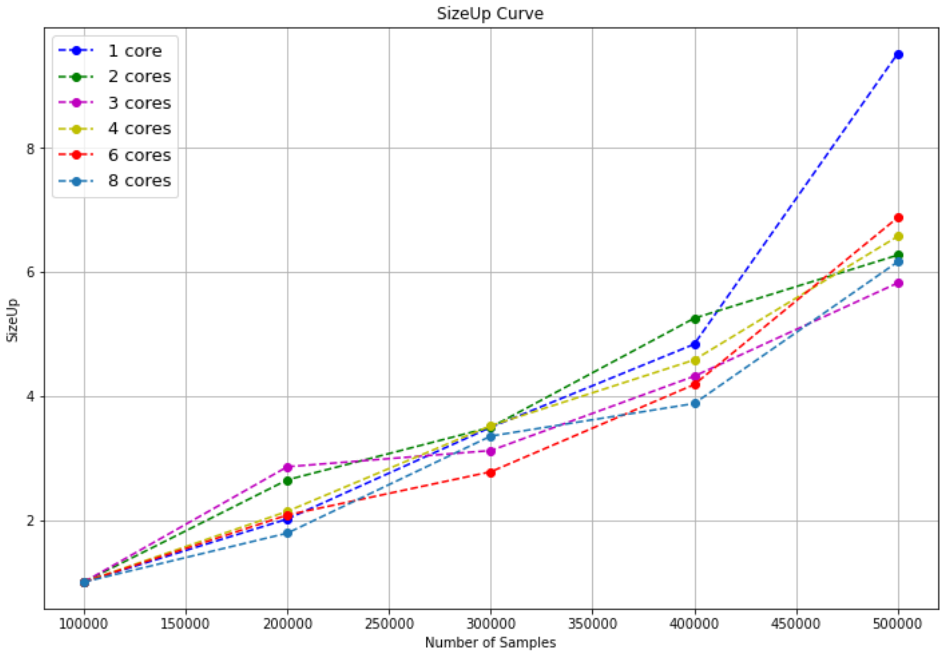
```
speedup_data
```

Out[30]:

	No. of Samples	Serial	2 Cores	3 Cores	4 Cores	6 Cores	8 Cores
0	100000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000
1	200000	2.016949	2.648437	2.859838	2.140647	2.081368	1.789030
2	300000	3.497030	3.496881	3.121142	3.522539	2.778994	3.355116
3	400000	4.837585	5.256197	4.321378	4.583883	4.186379	3.880265
4	500000	9.527135	6.275261	5.825042	6.577111	6.880631	6.166684

In [31]:

```
plt.figure(figsize=(12,8))
plt.plot(speedup_data['No. of Samples'],speedup_data['Serial'],'-b',linestyle = '--',marker='o', label='1 core')
plt.plot(speedup_data['No. of Samples'],speedup_data['2 Cores'],'-g',linestyle = '--',marker='o', label='2 cores')
plt.plot(speedup_data['No. of Samples'],speedup_data['3 Cores'],'-m',linestyle = '--',marker='o', label='3 cores')
plt.plot(speedup_data['No. of Samples'],speedup_data['4 Cores'],'-y',linestyle = '--',marker='o', label='4 cores')
plt.plot(speedup_data['No. of Samples'],speedup_data['6 Cores'],'-r',linestyle = '--',marker='o', label='6 cores')
plt.plot(speedup_data['No. of Samples'],speedup_data['8 Cores'],'-x',linestyle = '--',marker='o', label='8 cores')
plt.xlabel('Number of Samples')
plt.ylabel('SizeUp')
plt.title('SizeUp Curve')
plt.legend(loc='upper left',prop={'size': 13})
plt.grid()
plt.show()
```



ScaleUp Curve

In [11]:

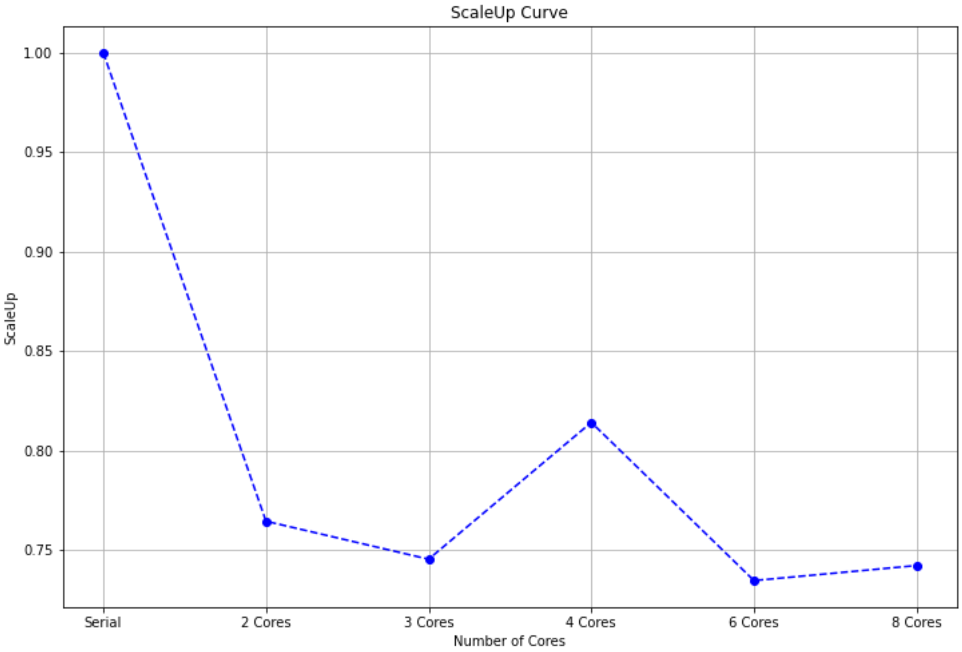
```
scaleup_data = pd.read_csv('Timing Data - ScaleUp.csv')
scaleup_data
```

Out[11]:

	No. of Cores	ScaleUp
0	Serial	1.000000
1	2 Cores	0.764403
2	3 Cores	0.745289
3	4 Cores	0.814053
4	6 Cores	0.734547
5	8 Cores	0.742111

In [12]:

```
plt.figure(figsize=(12,8))
plt.plot(scaleup_data['No. of Cores'],scaleup_data['ScaleUp'],'-b',linestyle = '--',marker='o')
plt.xlabel('Number of Cores')
plt.ylabel('ScaleUp')
plt.title('ScaleUp Curve')
plt.grid()
plt.show()
warnings.filterwarnings('ignore')
```


ScaleUp Curve
ScaleUp
Serial
2 Cores
3 Cores
4 Cores
6 Cores
8 Cores
Number of Cores