In [1]:
```
import pandas as pd
```

In [2]:
```
df = pd.read_csv('results/CLOUD_DATASIZE_1000.csv')
```

In [3]:
```
df['dataset_size'] = df['dataset_size']/100000000
```

In [4]:
```
df = df.round(2)
```

In [5]:
```
df
```

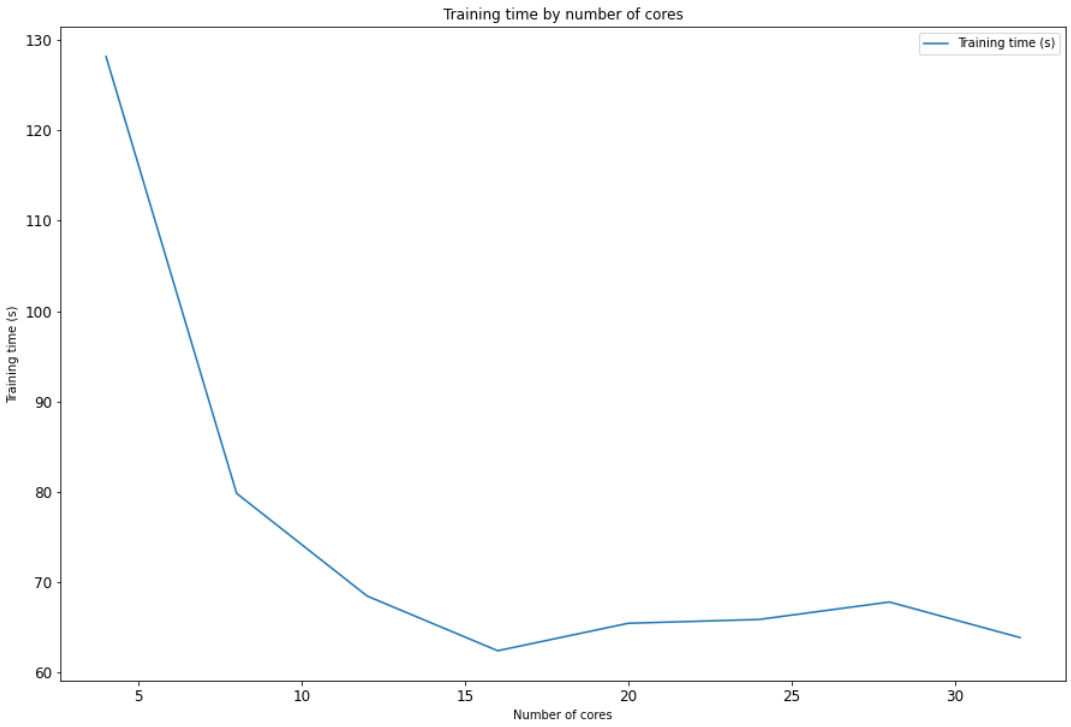
Out[5]:

	Unnamed: 0	time	dataset_rows	dataset_size_num	dataset_size	number_of_cores
0	0	128.16	32561000	1000	36.06	4
1	1	79.85	32561000	1000	36.06	8
2	2	68.49	32561000	1000	36.06	12
3	3	62.44	32561000	1000	36.06	16
4	4	65.48	32561000	1000	36.06	20
5	5	65.90	32561000	1000	36.06	24
6	6	67.83	32561000	1000	36.06	28
7	7	63.90	32561000	1000	36.06	32

In [6]:
```
df = df.round(2)
```

In [7]:
```
dftemp = df[['number_of_cores', 'time', ]]
dftemp.columns=['Number of cores', 'Training time (s)']
dftemp.plot.line(
    x='Number of cores',
    xlabel="Number of cores",
    ylabel="Training time (s)",
    rot=0,
    title='Training time by number of cores',
    figsize=(15,10),
    fontsize=12)
```

Out[7]:
```
<AxesSubplot:title={'center':'Training time by number of cores'}, xlabel='Number of cores', ylabel='Training time (s)'>
```



In [8]:
```
dftemp = df[['number_of_cores', 'dataset_size', 'time', ]]
dftemp.columns=['Number of cores', 'Dataset Size (GB)', 'Training time (s)']
dftemp
```

Out[8]:

	Number of cores	Dataset Size (GB)	Training time (s)
0	4	36.06	128.16
1	8	36.06	79.85
2	12	36.06	68.49
3	16	36.06	62.44
4	20	36.06	65.48
5	24	36.06	65.90
6	28	36.06	67.83
7	32	36.06	63.90

In [ ]: