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

In [2]:
```
df = pd.read_csv('results/CLOUD_NUMBER_OF_CORES_32.csv')
```

In [3]:
```
df = df.drop(columns=['Unnamed: 0'])
```

In [4]:
```
df['sf_train_time'] = df['train_time']/9.361921
```

In [5]:
```
df['sf_dataset_size'] = df['dataset_size']/360643033
```

In [6]:
```
df['dataset_size'] = df['dataset_size']/100000000
```

In [7]:
```
df['sf_difference'] = df['sf_dataset_size'] - df['sf_train_time']
```

In [8]:
```
df = df.round(2)
```

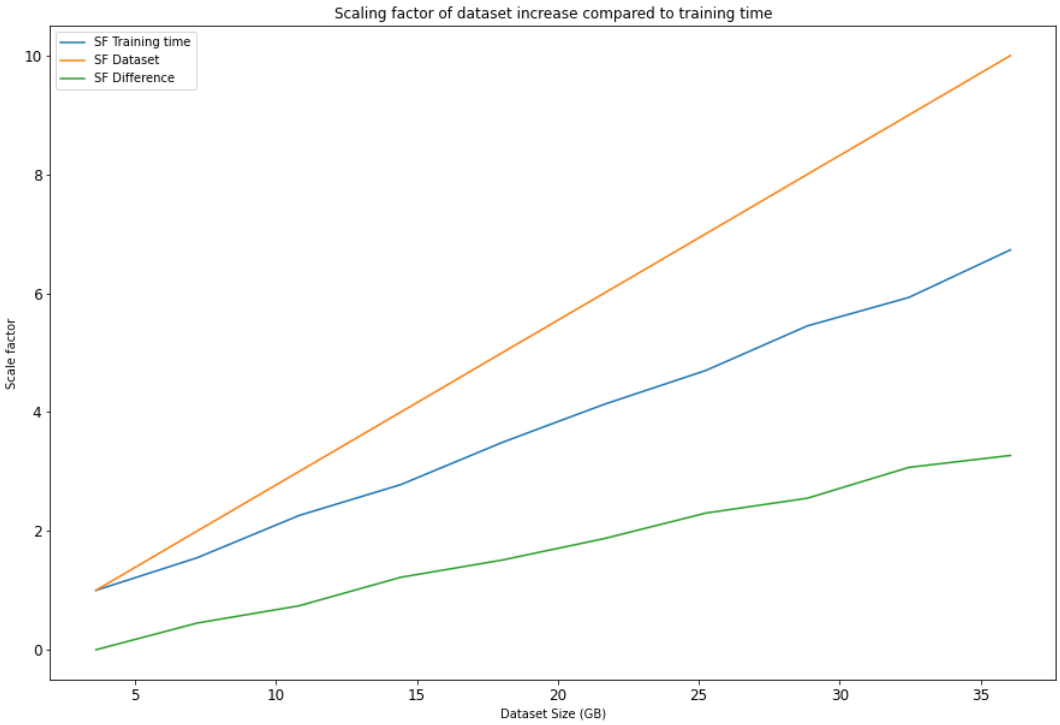
In [9]:
```
df
```

Out[9]:

	train_time	predict_time	time	dataset_rows	dataset_size_num	dataset_size	number_of_cores	sf_train_time	sf_dataset_size	sf_difference
0	9.36	0	9.36	3256100	100	3.61	32	1.00	1.0	0.00
1	14.47	0	14.47	6512200	200	7.21	32	1.55	2.0	0.45
2	21.15	0	21.15	9768300	300	10.82	32	2.26	3.0	0.74
3	26.03	0	26.03	13024400	400	14.43	32	2.78	4.0	1.22
4	32.69	0	32.69	16280500	500	18.03	32	3.49	5.0	1.51
5	38.64	0	38.64	19536600	600	21.64	32	4.13	6.0	1.87
6	43.96	0	43.96	22792700	700	25.25	32	4.70	7.0	2.30
7	51.05	0	51.05	26048800	800	28.85	32	5.45	8.0	2.55
8	55.54	0	55.54	29304900	900	32.46	32	5.93	9.0	3.07
9	62.97	0	62.97	32561000	1000	36.06	32	6.73	10.0	3.27

In [10]:
```
dftemp = df[['sf_train_time', 'sf_dataset_size', 'sf_difference', 'dataset_size']]
dftemp.columns=['SF Training time', 'SF Dataset', 'SF Difference', 'Dataset Size (GB)']
dftemp.plot.line(
    x='Dataset Size (GB)',
    xlabel="Dataset Size (GB)",
    ylabel="Scale factor",
    rot=0,
    title='Scaling factor of dataset increase compared to training time',
    figsize=(15,10),
    fontsize=12)
```

Out[10]:
```
<AxesSubplot:title={'center':'Scaling factor of dataset increase compared to training time'}, xlabel='Dataset Size (GB)', ylabel='Scale facto
r'>
```



In [11]:
```
dftemp = df[['dataset_size', 'time', 'sf_dataset_size', 'sf_train_time', 'sf_difference', ]]
dftemp.columns=['Dataset Size (GB)', 'Training time (s)', 'SF Dataset', 'SF Training time', 'SF Difference', ]
dftemp
```

Out[11]:

	Dataset Size (GB)	Training time (s)	SF Dataset	SF Training time	SF Difference
0	3.61	9.36	1.0	1.00	0.00
1	7.21	14.47	2.0	1.55	0.45
2	10.82	21.15	3.0	2.26	0.74
3	14.43	26.03	4.0	2.78	1.22
4	18.03	32.69	5.0	3.49	1.51
5	21.64	38.64	6.0	4.13	1.87
6	25.25	43.96	7.0	4.70	2.30
7	28.85	51.05	8.0	5.45	2.55
8	32.46	55.54	9.0	5.93	3.07
9	36.06	62.97	10.0	6.73	3.27

In [ ]: