**Practical - 9**

**Implementation of K Mean Clustering and Unclustering on Jupyter Notebook using Python.**

In [1]:

import matplotlib.pyplot as plt

import pandas as pd

import seaborn as sns

In [2]:

iris = sns.load\_dataset('iris')

labels = iris.species.unique()

In [3]:

iris.head()

Out[3]:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **sepal\_length** | **sepal\_width** | **petal\_length** | **petal\_width** | **species** |
| **0** | 5.1 | 3.5 | 1.4 | 0.2 | setosa |
| **1** | 4.9 | 3.0 | 1.4 | 0.2 | setosa |
| **2** | 4.7 | 3.2 | 1.3 | 0.2 | setosa |
| **3** | 4.6 | 3.1 | 1.5 | 0.2 | setosa |
| **4** | 5.0 | 3.6 | 1.4 | 0.2 | setosa |

In [4]:

iris["species"] = pd.Categorical(iris["species"])

iris["species"] = iris["species"].cat.codes

iris.head()

Out[4]:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **sepal\_length** | **sepal\_width** | **petal\_length** | **petal\_width** | **species** |
| **0** | 5.1 | 3.5 | 1.4 | 0.2 | 0 |
| **1** | 4.9 | 3.0 | 1.4 | 0.2 | 0 |
| **2** | 4.7 | 3.2 | 1.3 | 0.2 | 0 |
| **3** | 4.6 | 3.1 | 1.5 | 0.2 | 0 |
| **4** | 5.0 | 3.6 | 1.4 | 0.2 | 0 |

In [5]:

X = iris[['sepal\_length','sepal\_width']].values

y = iris.species

In [8]:

from sklearn.cluster import KMeans

model = KMeans(n\_clusters = 3).fit(X)

centers = model.cluster\_centers\_

new\_labels = model.labels\_

In [9]:

print('Centroids :',centers)

print('\nLabels :',new\_labels)

Centroids : [[5.77358491 2.69245283]

[6.81276596 3.07446809]

[5.006 3.428 ]]

Labels : [2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 1 1 0 1 0 1 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 1 1 1 1 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 1 1 1 1 0 1 1 1 1 1 1 0 0 1 1 1 1 0 1 0 1 0 1 1 0 0 1 1 1 1 1 0 0 1 1 1 0 1 1 1 0 1 1 1 0 1 1 0]

In [10]:

plt.figure(figsize=(12,8))

plt.scatter(X[:, 0], X[:, 1],c=y, s=60)

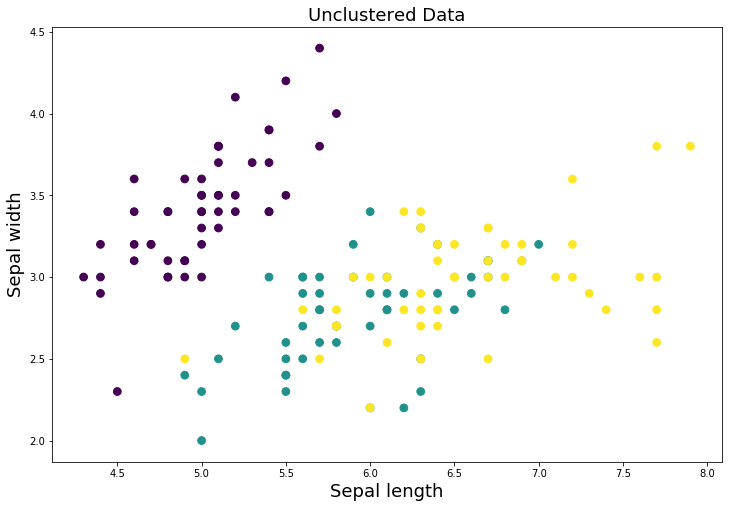
plt.xlabel('Sepal length', fontsize=18)

plt.ylabel('Sepal width', fontsize=18)

plt.title('Unclustered Data',fontsize=18)

Out[10]:

Text(0.5, 1.0, 'Unclustered Data')



In [12]:

plt.figure(figsize=(12,8))

plt.scatter(X[:, 0], X[:, 1], c=new\_labels,s=60)

plt.scatter(centers[:, 0], centers[:, 1], c='r', s=400, marker = '\*', zorder=10);

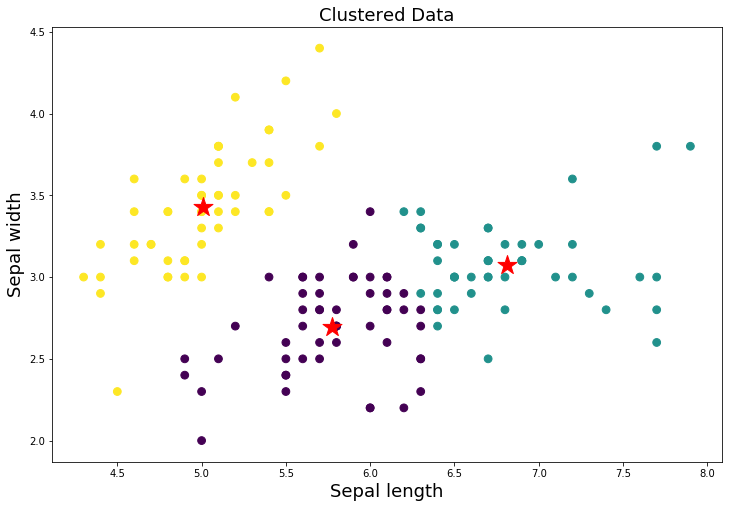
plt.xlabel('Sepal length', fontsize=18)

plt.ylabel('Sepal width', fontsize=18)

plt.title('Clustered Data',fontsize=18)

Out[12]:

Text(0.5, 1.0, 'Clustered Data')



In [15]:

y\_pred = model.predict([[2.3,5.6]])

print("Result :",labels[y\_pred[0]])

Result : virginica