Name : Mahmoud Hamed Ismael

Sparks Task 1

Predict the Percentage of an student based on study hours

In [2]:
```
#Importing the libraries


import pandas as pd
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split
import matplotlib.pyplot as plt
import seaborn as sns
sns.set()
```

In [26]:
```
#Importing the dataset

df = pd.read_csv('http://bit.ly/w-data')
X = df.iloc[:,:-1].values
y = df.iloc[:,-1].values
```

In [27]:
```
#Splitting the dataset into the Training set and Test set

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=0)
```

In [28]:
```
#Training the Simple Linear Regression model on the Training set

model = LinearRegression()
model.fit(X, y)
```

Out[28]:
```
LinearRegression()
```

In [29]:
```
#Predicting the Test set results

y_pred = model.predict(X_test)
```

In [30]:
```
#What will be predicted score if a student studies for 9.25 hrs/ day
float(model.predict([[9.25]]))
```

Out[30]:
```
92.90985477015731
```

In [32]:
```
#Visualising the Training set results
plt.scatter(X_train, y_train, color = 'r')
plt.plot(X_train, model.predict(X_train), color = 'b')
plt.title('Hours vs Scores (Training set)')
plt.xlabel('Hours')
plt.ylabel('Scores')
plt.show()
```

In [33]:
```
#Visualising the Test set results

plt.scatter(X_test, y_test, c= 'r')
plt.plot(X_test, model.predict(X_test))
plt.title('Hours vs Scores (Testing set)')
plt.xlabel('Hours')
plt.ylabel('Scores')
plt.show()
```