Practical 1

AIM :- Perform Linear regression on a data warehouse data.

Step 1 :- Go to RStudio -> Click on file -> New file -> go to Rscript

Step 2:-

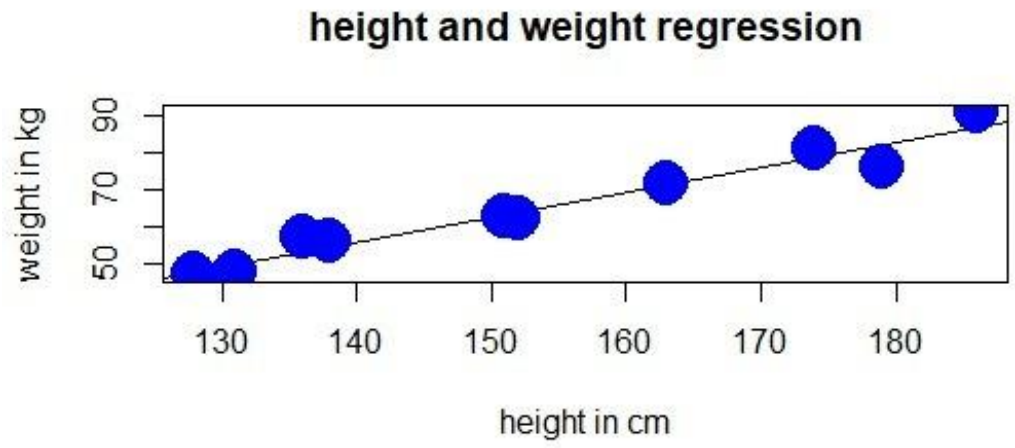
```
height <- c(151, 174, 138, 186, 128, 136, 179, 163, 152, 131)

weight <- c(63, 81, 56, 91, 47, 57, 76, 72, 62, 48)

height

weight

print(height)

relation<-lm(weight~height)

print(relation)

print(summary(relation))

a<-data.frame(height=170)

result<-predict(relation,a)

print(result)

plot(height,weight,col="blue",main="height and weight regression",

abline(relation),cex=3,pch=16,xlab="height in cm",ylab="weight in kg")
```

Command

```
> height <- c(151, 174, 138, 186, 128, 136, 179, 163, 152, 131)
> weight <- c(63, 81, 56, 91, 47, 57, 76, 72, 62, 48)
> height
 [1] 151 174 138 186 128 136 179 163 152 131
> weight
 [1] 63 81 56 91 47 57 76 72 62 48
> print(height)
 [1] 151 174 138 186 128 136 179 163 152 131
> relation<-lm(weight~height)
> print(relation)

Call:
lm(formula = weight ~ height)

Coefficients:
(Intercept)       height
   -38.4551       0.6746

> print(summary(relation))

Call:
lm(formula = weight ~ height)

Residuals:
    Min      1Q  Median      3Q     Max
-6.3002 -1.6629  0.0412  1.8944  3.9775

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -38.45509    8.04901  -4.778  0.00139 **
height        0.67461    0.05191  12.997 1.16e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.253 on 8 degrees of freedom
```

```
Multiple R-squared:  0.9548, Adjusted R-squared:  0.9491
F-statistic: 168.9 on 1 and 8 DF,  p-value: 1.164e-06

> a<-data.frame(height=170)
> result<-predict(relation,a)
> print(result)
       1
76.22869
> plot(height,weight,col="blue",main="height and weight regression",
+ abline(relation),cex=3,pch=16,xlab="height in cm",ylab="weight in kg")
```

OUTPUT:



Practical 2

AIM: Perform Linear regression on a data warehouse data

Step1:Go to R studio→create file→

Step2: Install catools → go to tools click on install package -> write package name then click on install.

```
quality=read.csv("C:\\Program Files\\RStudio\\quality_prac.csv")

str(quality)

summary(quality)

table(quality$PoorCare)

install.packages("caTools")

library(caTools)

set.seed(88)

split=sample.split(quality$PoorCare,SplitRatio = 0.75)

split

qualityTrain=subset(quality,split=TRUE)

qualityTest=subset(quality,split=FALSE)

numquality<-quality[,c(-1,-12)]

head(numquality)

cor(numquality)

Qualitylog=glm(PoorCare~OfficeVisits+Narcotics,data = qualityTrain,family = binomial)

summary(Qualitylog)

#make predictions on training set

predictTrain=predict(Qualitylog,type="response")

summary(predictTrain)

tapply(predictTrain, quality$PoorCare, mean)

table(quality$PoorCare,predictTrain>0.5)

install.packages("ROCR")

library(ROCR)

ROCRpred= prediction(predictTrain,quality$PoorCare)

ROCRperf=performance(ROCRpred,"tpr","fpr")

plot(ROCRperf,colorize=TRUE)
```

Command

```
Loading required package: ROCR

> quality=read.csv("C:\\Program Files\\RStudio\\quality_prac.csv")

> str(quality)
```

```
'data.frame':          131 obs. of  14 variables:
 $ MemberID          : int  1 2 3 4 5 6 7 8 9 10 ...
 $ InpatientDays     : int  0 1 0 0 8 2 16 2 2 4 ...
 $ ERVisits          : int  0 1 0 1 2 0 1 0 1 2 ...
 $ OfficeVisits      : int  18 6 5 19 19 9 8 8 4 0 ...
 $ Narcotics         : int  1 1 3 0 3 2 1 0 3 2 ...
 $ DaysSinceLastERVisit: num  731 411 731 158 449 ...
 $ Pain              : int  10 0 10 34 10 6 4 5 5 2 ...
 $ TotalVisits       : int  18 8 5 20 29 11 25 10 7 6 ...
 $ ProviderCount     : int  21 27 16 14 24 40 19 11 28 21 ...
 $ MedicalClaims     : int  93 19 27 59 51 53 40 28 20 17 ...
 $ ClaimLines        : int  222 115 148 242 204 156 261 87 98 66 ...
 $ StartedOnCombination: logi  FALSE FALSE FALSE FALSE FALSE FALSE ...
 $ AcuteDrugGapSmall   : int  0 1 5 0 0 4 0 0 0 0 ...
 $ PoorCare          : int  0 0 0 0 0 1 0 0 1 0 ...
> summary(quality)
   MemberID     InpatientDays      ERVisits      OfficeVisits
 Min.   :  1.0   Min.   : 0.000   Min.   : 0.000   Min.   : 0.00
 1st Qu.: 33.5   1st Qu.: 0.000   1st Qu.: 0.000   1st Qu.: 7.00
 Median : 66.0   Median : 0.000   Median : 1.000   Median :12.00
 Mean   : 66.0   Mean   : 2.718   Mean   : 1.496   Mean   :13.23
 3rd Qu.: 98.5   3rd Qu.: 3.000   3rd Qu.: 2.000   3rd Qu.:18.50
 Max.   :131.0   Max.   :30.000   Max.   :11.000   Max.   :46.00
  Narcotics     DaysSinceLastERVisit     Pain        TotalVisits
 Min.   : 0.000   Min.   :  6.0       Min.   :  0.00   Min.   : 0.00
 1st Qu.: 0.000   1st Qu.:207.0       1st Qu.:  1.00   1st Qu.: 8.00
 Median : 1.000   Median :641.0       Median :  8.00   Median :15.00
 Mean   : 4.573   Mean   :480.6       Mean   : 15.56   Mean   :17.44
 3rd Qu.: 3.000   3rd Qu.:731.0       3rd Qu.: 23.00   3rd Qu.:22.50
 Max.   :59.000   Max.   :731.0       Max.   :104.00   Max.   :69.00
 ProviderCount   MedicalClaims     ClaimLines    StartedOnCombination
 Min.   : 5.00   Min.   : 11.00   Min.   : 20.0   Mode :logical
 1st Qu.:15.00   1st Qu.: 25.50   1st Qu.: 83.5   FALSE:125
 Median :20.00   Median : 37.00   Median :120.0   TRUE :6
 Mean   :23.98   Mean   : 43.24   Mean   :142.9
```

```
 3rd Qu.:30.00   3rd Qu.: 49.50   3rd Qu.:185.0
 Max.   :82.00   Max.   :194.00   Max.   :577.0
 AcuteDrugGapSmall    PoorCare
 Min.   : 0.000    Min.   :0.0000
 1st Qu.: 0.000    1st Qu.:0.0000
 Median : 1.000    Median :0.0000
 Mean   : 2.695    Mean   :0.2519
 3rd Qu.: 3.000    3rd Qu.:0.5000
 Max.   :71.000    Max.   :1.0000
> table(quality$PoorCare)
 0  1
98 33
> install.packages("caTools")
> library(caTools)
> set.seed(88)
> split=sample.split(quality$PoorCare,SplitRatio = 0.75)
> split
  [1]  TRUE  TRUE  TRUE  TRUE FALSE  TRUE FALSE  TRUE FALSE FALSE  TRUE
 [12] FALSE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
 [23]  TRUE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE FALSE FALSE FALSE FALSE
 [34]  TRUE  TRUE  TRUE FALSE  TRUE  TRUE  TRUE FALSE FALSE  TRUE  TRUE
 [45] FALSE  TRUE FALSE  TRUE FALSE  TRUE  TRUE FALSE FALSE  TRUE  TRUE
 [56]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE
 [67]  TRUE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
 [78]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE
 [89]  TRUE  TRUE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE
[100]  TRUE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE  TRUE FALSE  TRUE FALSE
[111] FALSE  TRUE  TRUE FALSE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE FALSE
[122]  TRUE  TRUE FALSE  TRUE  TRUE FALSE  TRUE  TRUE  TRUE FALSE
> qualityTrain=subset(quality,split=TRUE)
> qualityTest=subset(quality,split=FALSE)
> numquality<-quality[,c(-1,-12)]
> head(numquality)
  InpatientDays ERVisits OfficeVisits Narcotics DaysSinceLastERVisit
1             0        0           18         1                  731
```

```
2          1      1         6       1             411

3          0      0         5       3             731

4          0      1        19       0             158

5          8      2        19       3             449

6          2      0         9       2             731

  Pain TotalVisits ProviderCount MedicalClaims ClaimLines

1   10        18          21         93       222

2    0         8          27         19       115

3   10         5          16         27       148

4   34        20          14         59       242

5   10        29          24         51       204

6    6        11          40         53       156

  AcuteDrugGapSmall PoorCare

1              0       0

2              1       0

3              5       0

4              0       0

5              0       0

6              4       1

> cor(numquality)

                 InpatientDays     ERVisits OfficeVisits

InpatientDays        1.000000000  0.440087299    0.1759011

ERVisits             0.440087299  1.000000000    0.3085257

OfficeVisits         0.175901119  0.308525685    1.0000000

Narcotics           -0.093768932 -0.003731653    0.2757593

DaysSinceLastERVisit  -0.290121046 -0.735246070   -0.1283879

Pain                 0.304058069  0.546779466    0.3529678

TotalVisits          0.622035618  0.586438628    0.8653868

ProviderCount        0.244023304  0.457429030    0.3654691

MedicalClaims        0.286377975  0.355318952    0.4985134

ClaimLines           0.386951074  0.542000500    0.4249532

AcuteDrugGapSmall    -0.001144346 -0.072749681    0.2007348

PoorCare             0.080725715  0.135400778    0.3295118

                  Narcotics DaysSinceLastERVisit       Pain

InpatientDays       -0.093768932        -0.29012105  0.30405807
```

ERVisits -0.003731653 -0.73524607 0.54677947

OfficeVisits 0.275759302 -0.12838788 0.35296784

Narcotics 1.000000000 0.06505481 0.10686036

DaysSinceLastERVisit 0.065054809 1.00000000 -0.35878080

Pain 0.106860359 -0.35878080 1.00000000

TotalVisits 0.163992449 -0.34463954 0.48295915

ProviderCount 0.293478180 -0.29770084 0.40509514

MedicalClaims 0.220540818 -0.19811441 0.29669718

ClaimLines 0.185798702 -0.41279666 0.46471274

AcuteDrugGapSmall 0.710888560 0.13108501 -0.03149016

PoorCare 0.447236064 -0.10798298 0.09216828

TotalVisits ProviderCount MedicalClaims

InpatientDays 0.6220356 0.2440233 0.2863780

ERVisits 0.5864386 0.4574290 0.3553190

OfficeVisits 0.8653868 0.3654691 0.4985134

Narcotics 0.1639924 0.2934782 0.2205408

DaysSinceLastERVisit -0.3446395 -0.2977008 -0.1981144

Pain 0.4829592 0.4050951 0.2966972

TotalVisits 1.0000000 0.4515455 0.5493080

ProviderCount 0.4515455 1.0000000 0.5170023

MedicalClaims 0.5493080 0.5170023 1.0000000

ClaimLines 0.5696186 0.6053573 0.8139345

AcuteDrugGapSmall 0.1348611 0.1412836 0.0856369

PoorCare 0.3005403 0.2201661 0.1673987

ClaimLines AcuteDrugGapSmall PoorCare

InpatientDays 0.38695107 -0.001144346 0.08072572

ERVisits 0.54200050 -0.072749681 0.13540078

OfficeVisits 0.42495323 0.200734789 0.32951181

Narcotics 0.18579870 0.710888560 0.44723606

DaysSinceLastERVisit -0.41279666 0.131085008 -0.10798298

Pain 0.46471274 -0.031490160 0.09216828

TotalVisits 0.56961864 0.134861075 0.30054033

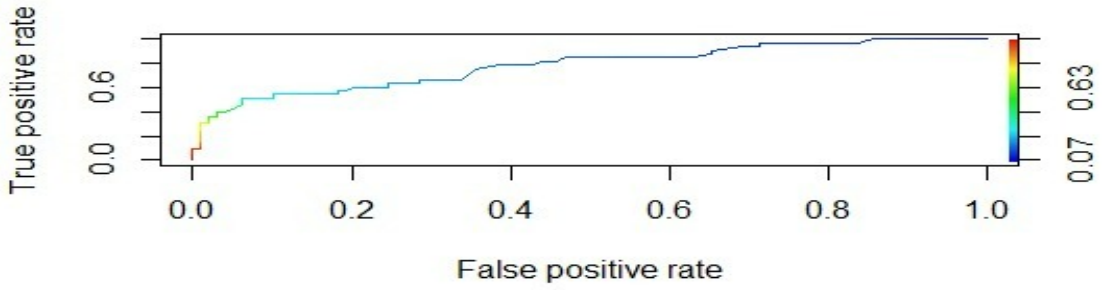
ProviderCount 0.60535725 0.141283618 0.22016613

MedicalClaims 0.81393452 0.085636905 0.16739875

ClaimLines 1.00000000 -0.013229464 0.12917477

```
AcuteDrugGapSmall    -0.01322946      1.000000000  0.34143466
PoorCare             0.12917477       0.341434658  1.00000000
> Qualitylog=glm(PoorCare~OfficeVisits+Narcotics,data = qualityTrain,family = binomial)
> summary(Qualitylog)
Call:
glm(formula = PoorCare ~ OfficeVisits + Narcotics, family = binomial,
   data = qualityTrain)
Deviance Residuals:
   Min      1Q   Median      3Q     Max
-2.3773  -0.6266  -0.5096  -0.1539   2.1186
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -2.54021    0.45005  -5.644 1.66e-08 *
OfficeVisits  0.06273    0.02399   2.615 0.008920 **
Narcotics     0.10990    0.03263   3.368 0.000757 *
Signif. codes:  0 ‘*’ 0.001 ‘*’ 0.01 ‘’ 0.05 ‘.’ 0.1 ‘ ’ 1
(Dispersion parameter for binomial family taken to be 1)
   Null deviance: 147.88  on 130  degrees of freedom
Residual deviance: 116.45  on 128  degrees of freedom
AIC: 122.45
Number of Fisher Scoring iterations: 5
> predictTrain=predict(Qualitylog,type="response")
> summary(predictTrain)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.07309 0.12179 0.15956 0.25191 0.26539 0.99667
> tapply(predictTrain, quality$PoorCare, mean)
      0       1
0.1856957 0.4485400
> table(quality$PoorCare,predictTrain>0.5)
   FALSE TRUE
 0   94    4
 1   20   13
```

OUTPUT:


True positive rate
0.0
0.6
0.0
0.2
0.4
0.6
0.8
1.0
False positive rate
0.07
0.63

Practical 3

AIM: Perform data classification using classification algorithm

Naive-bayes:-

Step1:Go to R studio→create file→

Step2: Install catools,e1071,caret → go to tools click on install package -> write package name then click on install

Step3:

```
#classification using NAIVE BAYES THEOREM
data("iris")
str(iris)
install.packages("e1071") #predict function
install.packages("caTools")
install.packages("caret")
library(e1071)
library(caTools)
library(caret)
split<-sample.split(iris,SplitRatio = 0.7)
split
train_cl<-subset(iris,split=TRUE)
test_cl<-subset(iris,split=FALSE)
View(iris)
train_scale<-scale(train_cl[,1:4])
test_scale<-scale(test_cl[,1:4])
View(train_scale)
set.seed(120)
classifier_cl<-naiveBayes(Species~.,data=train_cl)
y_pred<-predict(classifier_cl,test_cl)
y_pred
cm<-table(test_cl$Species,y_pred)
cm
confusionMatrix(cm)
> #classification using NAIVE BAYES THEOREM
```

```
> data("iris")
> str(iris)
'data.frame':	150 obs. of  5 variables:
 $ Sepal.Length: num  5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...
 $ Sepal.Width : num  3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...
 $ Petal.Length: num  1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...
 $ Petal.Width : num  0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...
 $ Species     : Factor w/ 3 levels "setosa","versicolor",..: 1 1 1 1 1 1 1 1 1 1 ...
> install.packages("e1071") #predict function
package 'e1071' successfully unpacked and MD5 sums checked
The downloaded binary packages are in
	C:\Users\Lenovo\AppData\Local\Temp\RtmpSEozMT\downloaded_packages
> install.packages("caTools")
> install.packages("caret")
package 'caret' successfully unpacked and MD5 sums checked
The downloaded binary packages are in
	C:\Users\Lenovo\AppData\Local\Temp\RtmpSEozMT\downloaded_packages
> library(e1071)
> library(caTools)
> library(caret)
Loading required package: ggplot2
Want to understand how all the pieces fit together? Read R for
Data Science: https://r4ds.had.co.nz/
Loading required package: lattice
> split<-sample.split(iris,SplitRatio = 0.7)
> split
[1]  TRUE  TRUE  TRUE FALSE FALSE
> train_cl<-subset(iris,split=TRUE)
> test_cl<-subset(iris,split=FALSE)
> View(iris)
> train_scale<-scale(train_cl[,1:4])
> test_scale<-scale(test_cl[,1:4])
> View(train_scale)
> set.seed(120)
> classifier_cl<-naiveBayes(Species~.,data=train_cl)
```

```
> y_pred<-predict(classifier_cl,test_cl)
> y_pred
 [1] setosa     setosa     setosa     setosa     setosa     setosa
[7] setosa     setosa     setosa     setosa     setosa     setosa
 [13] setosa     setosa     setosa     setosa     setosa     setosa
 [19] setosa     setosa     setosa     setosa     setosa     setosa
 [25] setosa     setosa     setosa     setosa     setosa     setosa
 [31] setosa     setosa     setosa     setosa     setosa     setosa
 [37] setosa     setosa     setosa     setosa     setosa     setosa
 [43] setosa     setosa     setosa     setosa     setosa     setosa
 [49] setosa     setosa     versicolor versicolor virginica  versicolor
 [55] versicolor versicolor versicolor versicolor versicolor versicolor
 [61] versicolor versicolor versicolor versicolor versicolor versicolor
 [67] versicolor versicolor versicolor versicolor virginica  versicolor
 [73] versicolor versicolor versicolor versicolor versicolor virginica
 [79] versicolor versicolor versicolor versicolor versicolor versicolor
 [85] versicolor versicolor versicolor versicolor versicolor versicolor
 [91] versicolor versicolor versicolor versicolor versicolor versicolor
 [97] versicolor versicolor versicolor versicolor virginica  virginica
[103] virginica  virginica  virginica  virginica  versicolor virginica
[109] virginica  virginica  virginica  virginica  virginica  virginica
[115] virginica  virginica  virginica  virginica  virginica  versicolor
[121] virginica  virginica  virginica  virginica  virginica  virginica
[127] virginica  virginica  virginica  virginica  virginica  virginica
[133] virginica  versicolor virginica  virginica  virginica  virginica
[139] virginica  virginica  virginica  virginica  virginica  virginica
[145] virginica  virginica  virginica  virginica  virginica  virginica
Levels: setosa versicolor virginica
> cm<-table(test_cl$Species,y_pred)
> cm
            y_pred
             setosa versicolor virginica
  setosa         50          0         0
  versicolor      0         47         3
  virginica       0          3        47
```

> confusionMatrix(cm)

Confusion Matrix and Statistic

y_pred

setosa versicolor virginica

setosa 50 0 0

versicolor 0 47 3

virginica 0 3 47

Overall Statistics

Accuracy : 0.96

95% CI : (0.915, 0.9852)

No Information Rate : 0.3333

P-Value [Acc > NIR] : < 2.2e-16

Kappa : 0.94

Mcnemar's Test P-Value : NA

Statistics by Class

Class: setosa Class: versicolor Class: virginica

Sensitivity 1.0000 0.9400 0.9400

Specificity 1.0000 0.9700 0.9700

Pos Pred Value 1.0000 0.9400 0.9400

Neg Pred Value 1.0000 0.9700 0.9700

Prevalence 0.3333 0.3333 0.3333

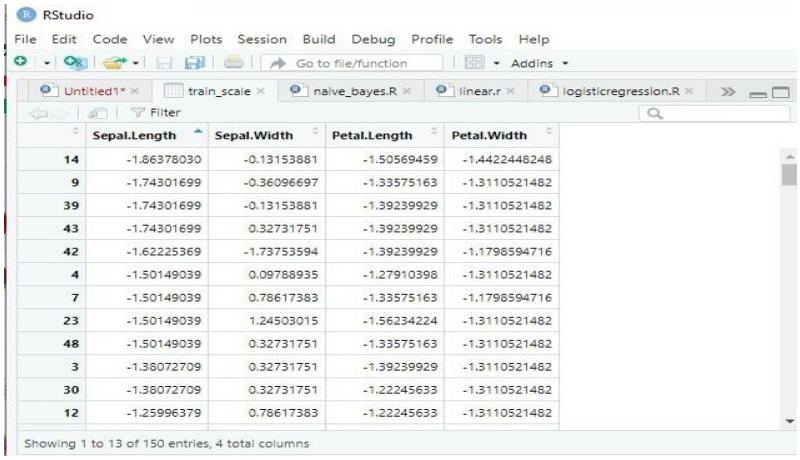
Detection Rate 0.3333 0.3133 0.3133

Detection Prevalence 0.3333 0.3333 0.3333

Balanced Accuracy 1.0000 0.9550 0.9550

> install.packages("caTools")

OUTPUT:



Practical 4

AIM :- Data clustering using clustering algorithm.

Kmeans cluster :-

Step 1 :- Go to RStudio -> Click on file -> New file -> go to Rscript

Step 2:- Install ggplot2,cluster :- go to tools click on install package -> write package name then click on install.

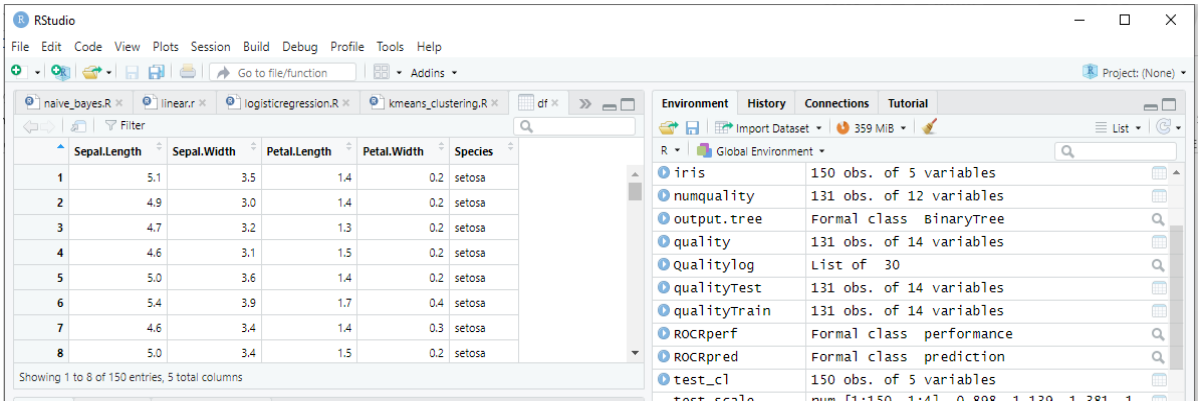
Step 3:-

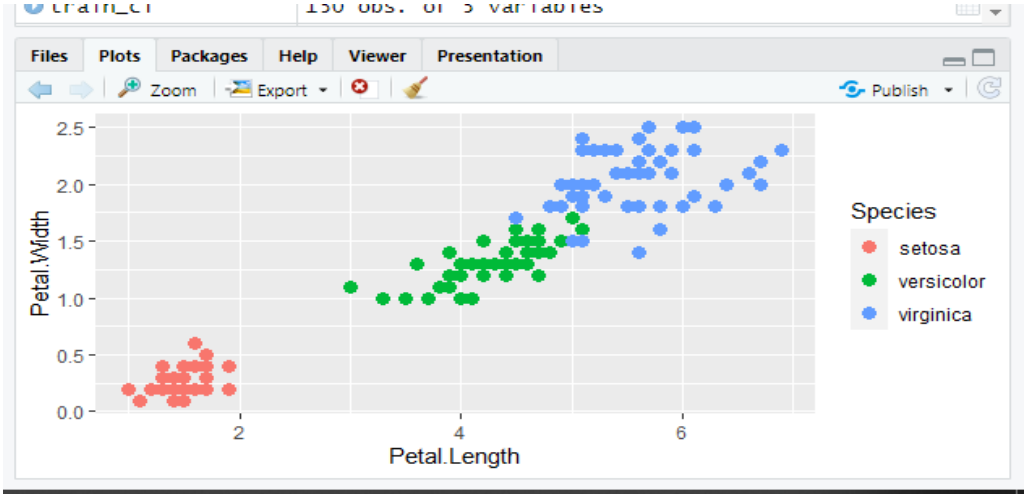
```
install.packages("ggplot2")
install.packages("cluster")
library(ggplot2)
library(cluster)
data(iris)
df<-iris
View(df)
head(df)
ggplot(df,aes(Petal.Length,Petal.Width))+
geom_point(aes(col=Species),size=3)
set.seed(101)
irisCluster<-kmeans(df[,1:4],centers = 3)
irisCluster
clusplot(iris,irisCluster$cluster,color = T,
shade = T,labels = 1,lines = 2
> install.packages("ggplot2")
 (as ‘lib’ is unspecified)
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.2/ggplot2_3.4.1.zip'
Content type 'application/zip' length 4226509 bytes (4.0 MB)
downloaded 4.0 MB
package ‘ggplot2’ successfully unpacked and MD5 sums checked
The downloaded binary packages are in
        C:\Users\Lenovo\AppData\Local\Temp\RtmpQlrTrC\downloaded_packages
> install.packages("cluster")
 (as ‘lib’ is unspecified)
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.2/cluster_2.1.4.zip'
Content type 'application/zip' length 585850 bytes (572 KB)
downloaded 572 KB
```

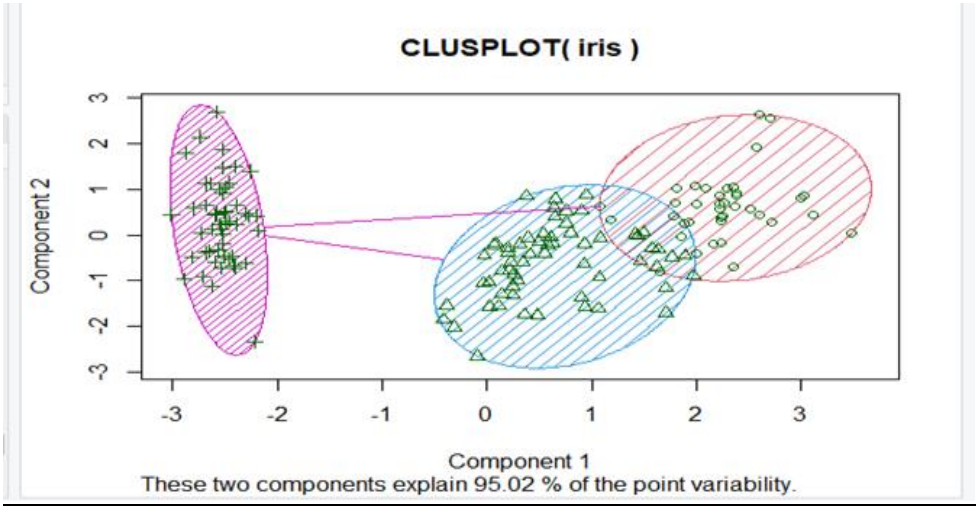
```
package ‘cluster’ successfully unpacked and MD5 sums checked
The downloaded binary packages are in
        C:\Users\Lenovo\AppData\Local\Temp\RtmpQlrTrC\downloaded_packages
> library(ggplot2)
> library(cluster)
> library(cluster)
> data(iris)
> install.packages("ggplot2")
> data(iris)
Loading required package: ROCR
 (as ‘lib’ is unspecified)
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.2/ggplot2_3.4.1.zip'
Content type 'application/zip' length 4226509 bytes (4.0 MB)
downloaded 4.0 MB
package ‘ggplot2’ successfully unpacked and MD5 sums checked
The downloaded binary packages are in
        C:\Users\Lenovo\AppData\Local\Temp\RtmpOSauqp\downloaded_packages
> install.packages("ggplot2")
 (as ‘lib’ is unspecified)
library(ggplot2)
> library(cluster)
> data(iris)
> df<-iris
> View(df)
> head(df)
  Sepal.Length Sepal.Width Petal.Length Petal.Width Species
1          5.1         3.5          1.4         0.2  setosa
2          4.9         3.0          1.4         0.2  setosa
3          4.7         3.2          1.3         0.2  setosa
4          4.6         3.1          1.5         0.2  setosa
5          5.0         3.6          1.4         0.2  setosa
6          5.4         3.9          1.7         0.4  setosa
> ggplot(df,aes(Petal.Length,Petal.Width))+
+ geom_point(aes(col=Species),size=3)
> set.seed(101)
```

```
> irisCluster<-kmeans(df[,1:4],centers = 3)

> irisCluster

K-means clustering with 3 clusters of sizes 38, 62, 50

Cluster means:

  Sepal.Length Sepal.Width Petal.Length Petal.Width

1    6.850000    3.073684    5.742105    2.071053

2    5.901613    2.748387    4.393548    1.433871

3    5.006000    3.428000    1.462000    0.246000

Clustering vector:

 [1] 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3

 [34] 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 2 2 1 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

 [67] 2 2 2 2 2 2 2 2 2 2 2 1 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

[100] 2 1 2 1 1 1 1 2 1 1 1 1 1 1 2 2 1 1 1 1 2 1 2 1 2 1 1 2 2 1 1 1 1

[133] 1 2 1 1 1 1 2 1 1 1 2 1 1 1 2 1 1 2

Within cluster sum of squares by cluster:

[1] 23.87947 39.82097 15.15100

 (between_SS / total_SS =  88.4 %)

Available components:

[1] "cluster"     "centers"     "totss"       "withinss"

[5] "tot.withinss" "betweenss"   "size"        "iter"

[9] "ifault"

> clusplot(iris,irisCluster$cluster,color = T,

+ shade = T,labels = 1,lines = 2)
```

OUTPUT:




Files
Plots
Packages
Help
Viewer
Presentation
Zoom
Export
Publish
Petal.Width
Petal.Length
Species
setosa
versicolor
virginica


CLUSPLOT(iris)
Component 2
Component 1
These two components explain 95.02 % of the point variability.

Practical 5

AIM:-Perform ETL process to construct in Power BI

Step 1:- Go to Power BI software -> Click on Get Data -> Select OData feed.

ETL Process in Power BI

1) Remove other columns to only display columns of interest

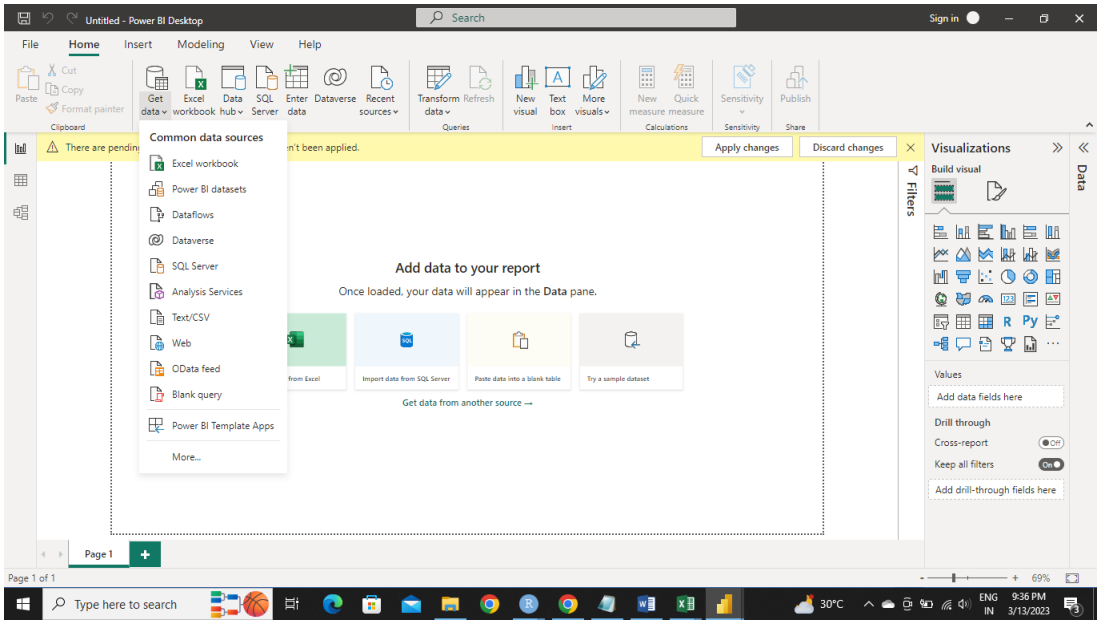
In this step you remove all columns except ProductID, ProductName, UnitsInStock, and QuantityPerUnit

Power BI Desktop includes Query Editor, which is where you shape and transform your data connections.

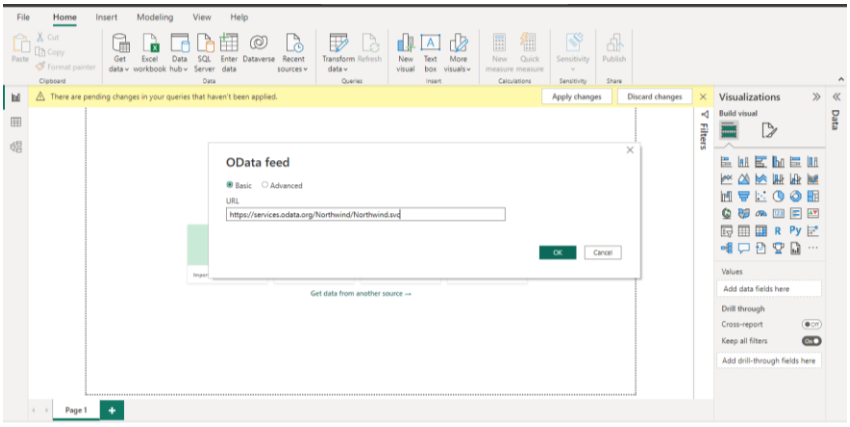
Query Editor opens automatically when you select **Edit** from Navigator. You can also open the Query Editor by selecting Edit Queries from the Home ribbon in Power BI Desktop. The following steps are performed in Query Editor.

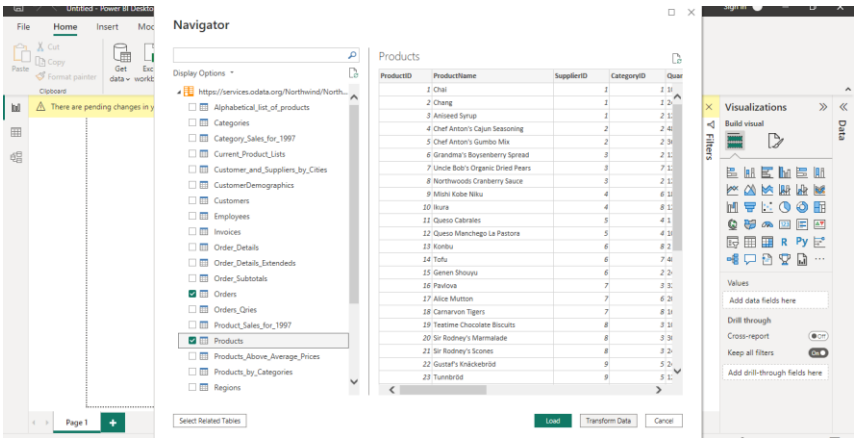
- In **Query Editor**, select the **ProductID, ProductName**, **QuantityPerUnit,** and **UnitsInStock** columns (use **Ctrl+Click** to select more than one column, or **Shift+Click** to select columns that are beside each other).

Select **Remove Columns** > **Remove** Other Columns from the ribbon, or right-click on a column header and click Remove Other Columns.

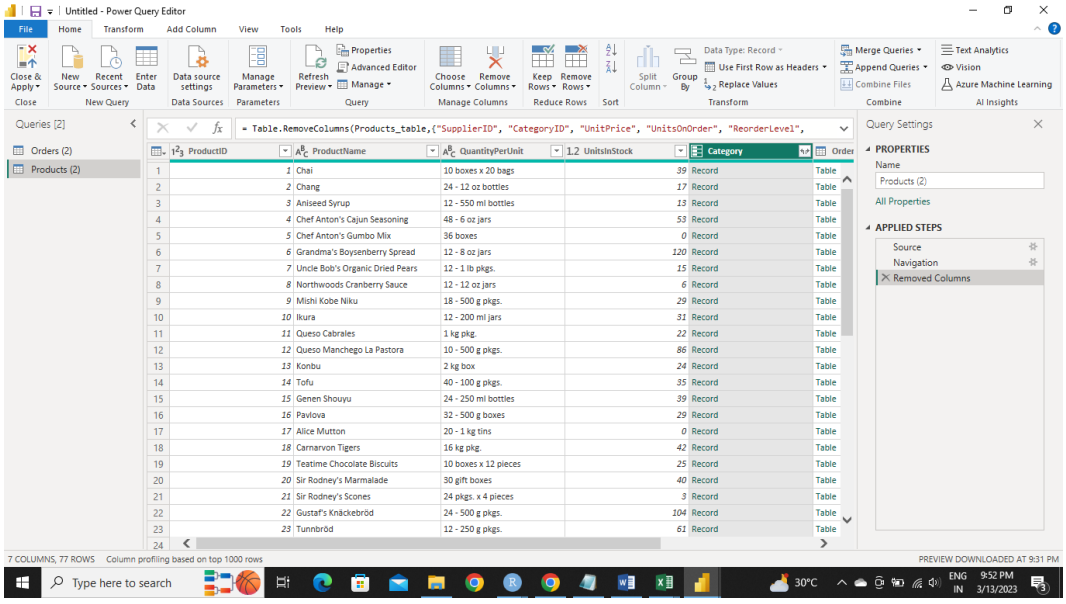


Step 2:-Put the URL -> click on ok -> Navigator window select Table :- Order and Product -> click on Transform Data

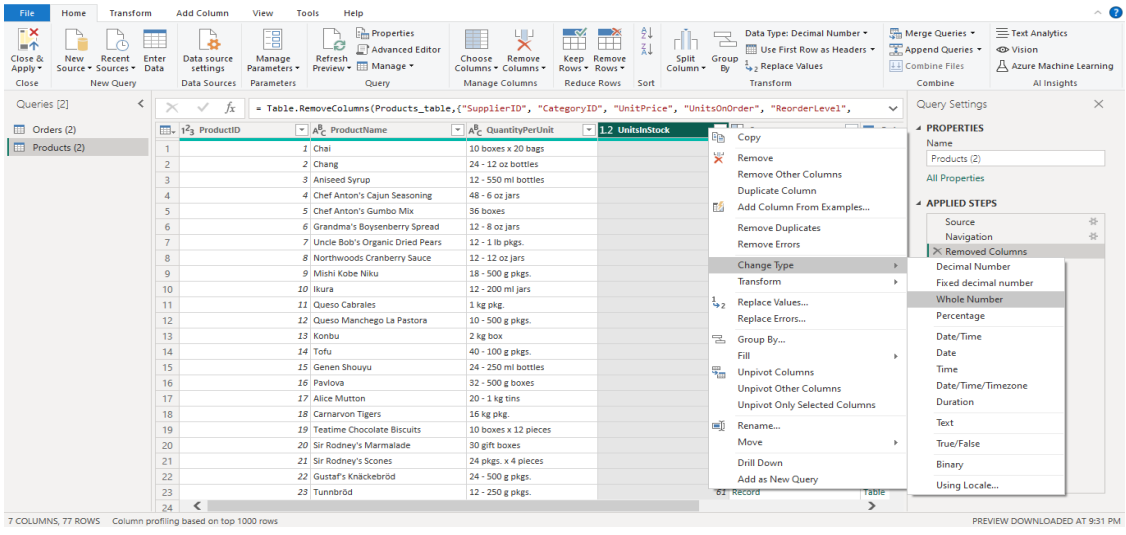




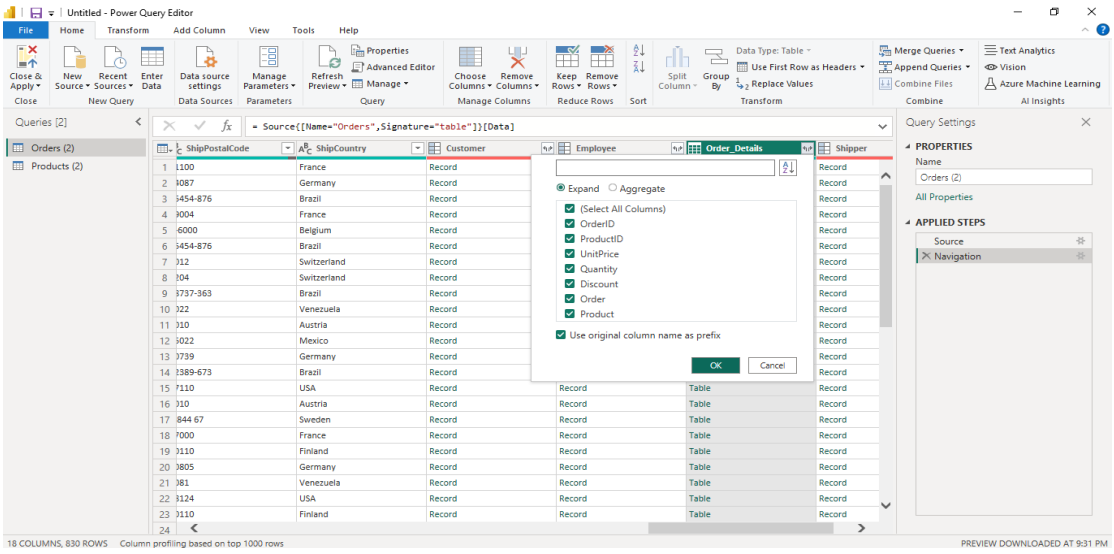
Step 3 :- Remove unused column



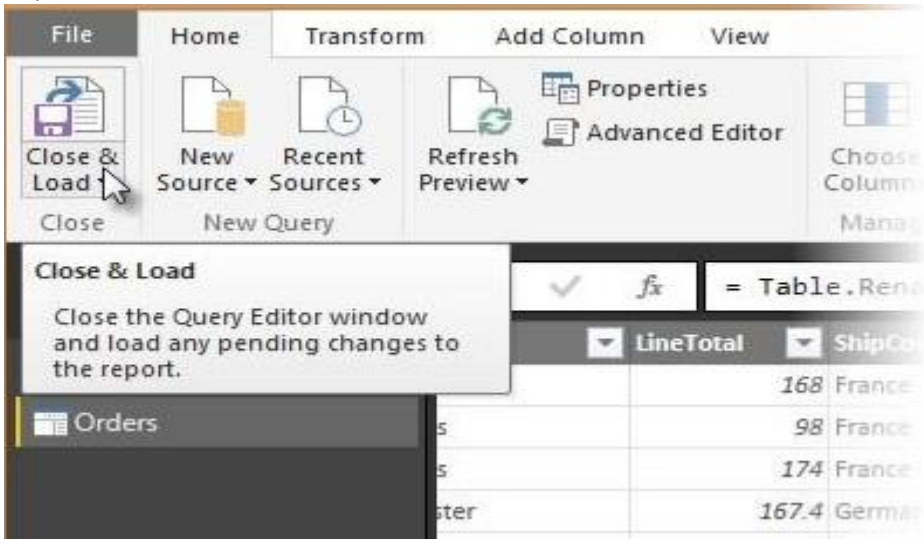
Step 4 :- Change the data type of UnitsInStock -> Right click on UnitsInStock column -> Select change Type -> Select whole number



Step 5 :- Go to Order Table -> go to order detail column -> click on right side of order detail column .-> select which column you want

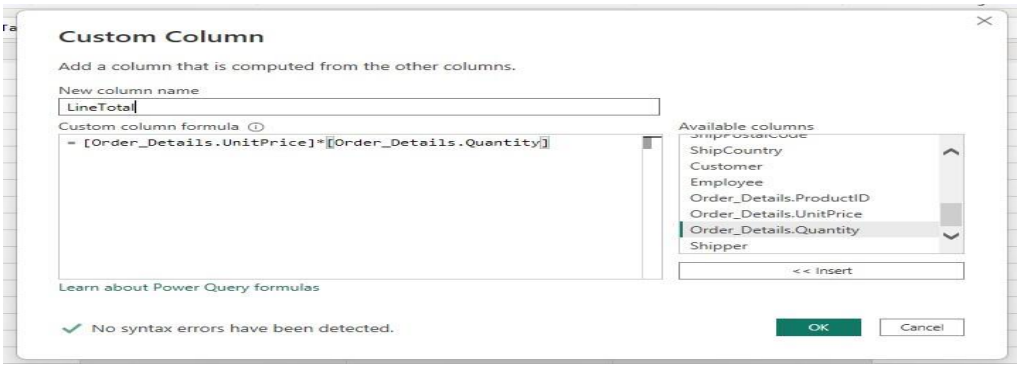


Step 6:- Calculate the line total for each Order Details row



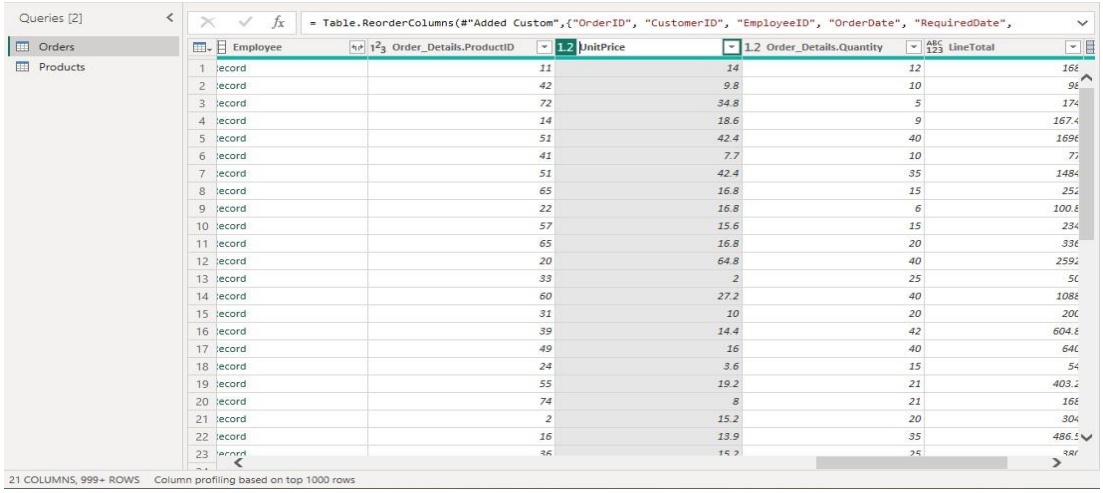
1. In the Add Column ribbon tab, click Add Custom Column.
2. In the Add Custom Column dialog box, in the Custom Column Formula textbox, enter [Order_Details.UnitPrice] * [Order_Details.Quantity].

3. In the New column name textbox, enter LineTotal.
4. Click OK.



Step 5:- Rename and reorder columns in the query

1. In Query Editor, drag the LineTotal column to the left, after ShipCountry.



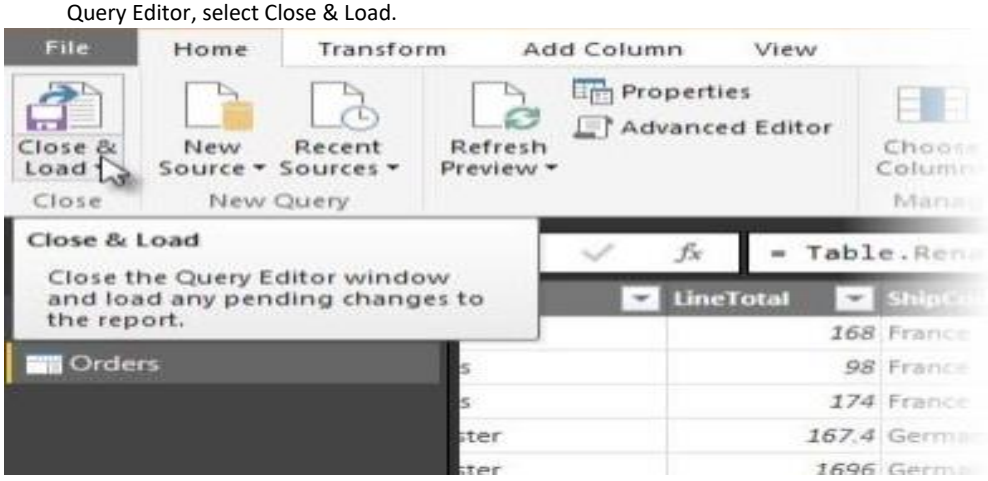
2. Remove the Order Details. Prefix from the Order_Details.ProductID, Order_Details.UnitPrice and Order_Details.Quantity columns, by double-clicking on each column header, and then deleting that text from the column name.

6. Combine the Products and Total Sales queries

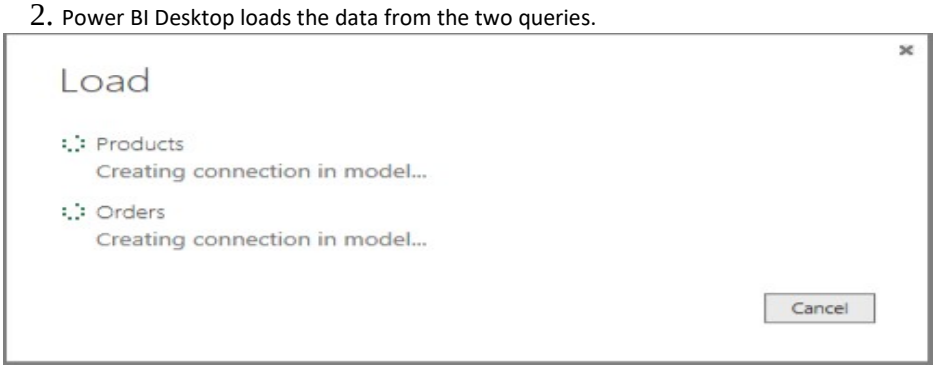
Power BI Desktop does not require you to combine queries to report on them. Instead, you can create Relationships between datasets. These relationships can be created on any column that is common to your datasets

Step 1: Confirm the relationship between Products and Total Sales

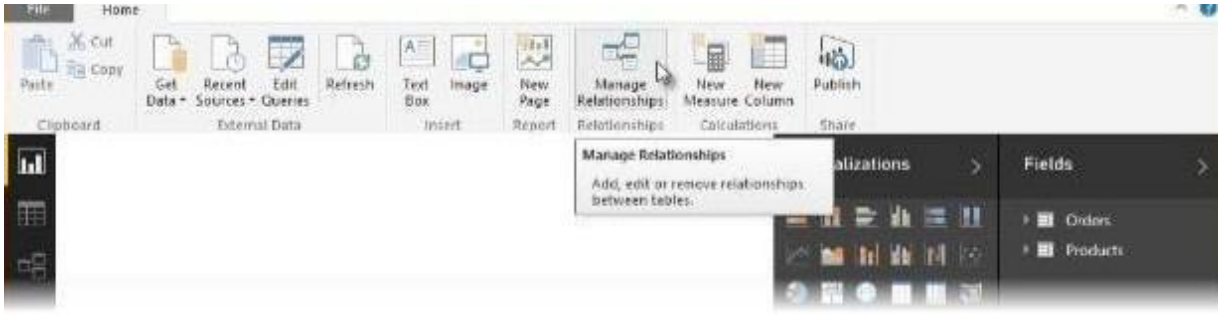
1. First, we need to load the model that we created in Query Editor into Power BI Desktop. From the Home ribbon of Query Editor, select Close & Load.



2. Power BI Desktop loads the data from the two queries.

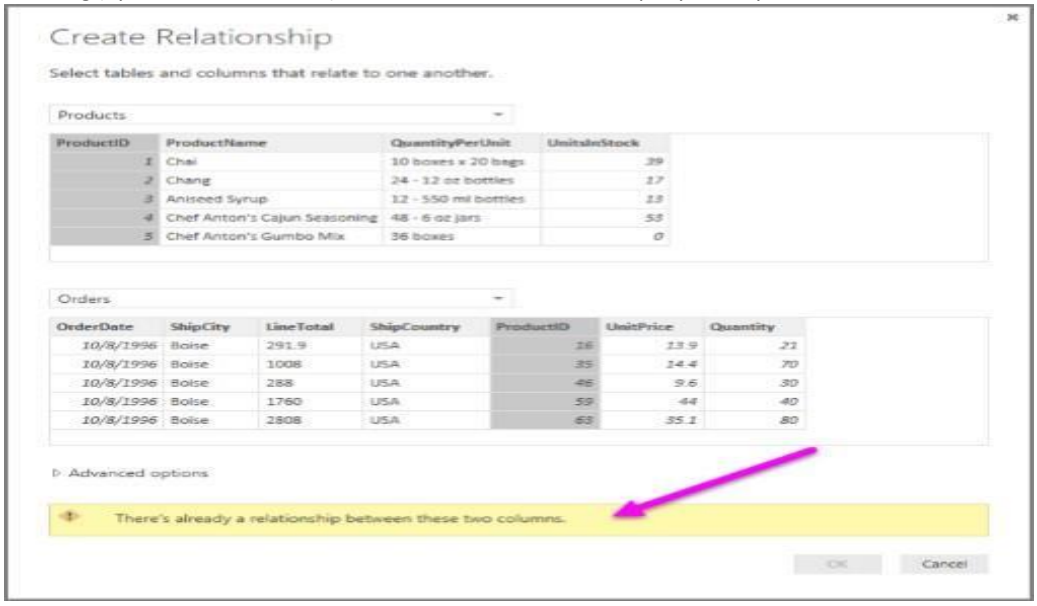


3. Once the data is loaded, select the Manage Relationships button Home ribbon.

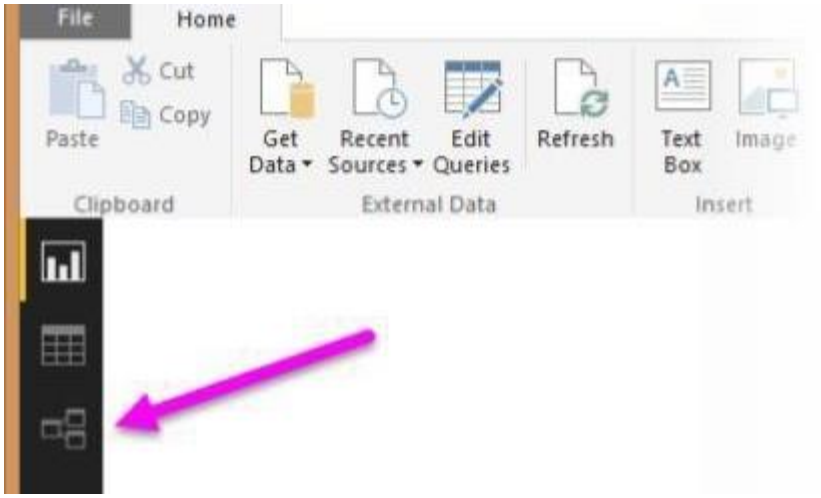


4. Select the New… button

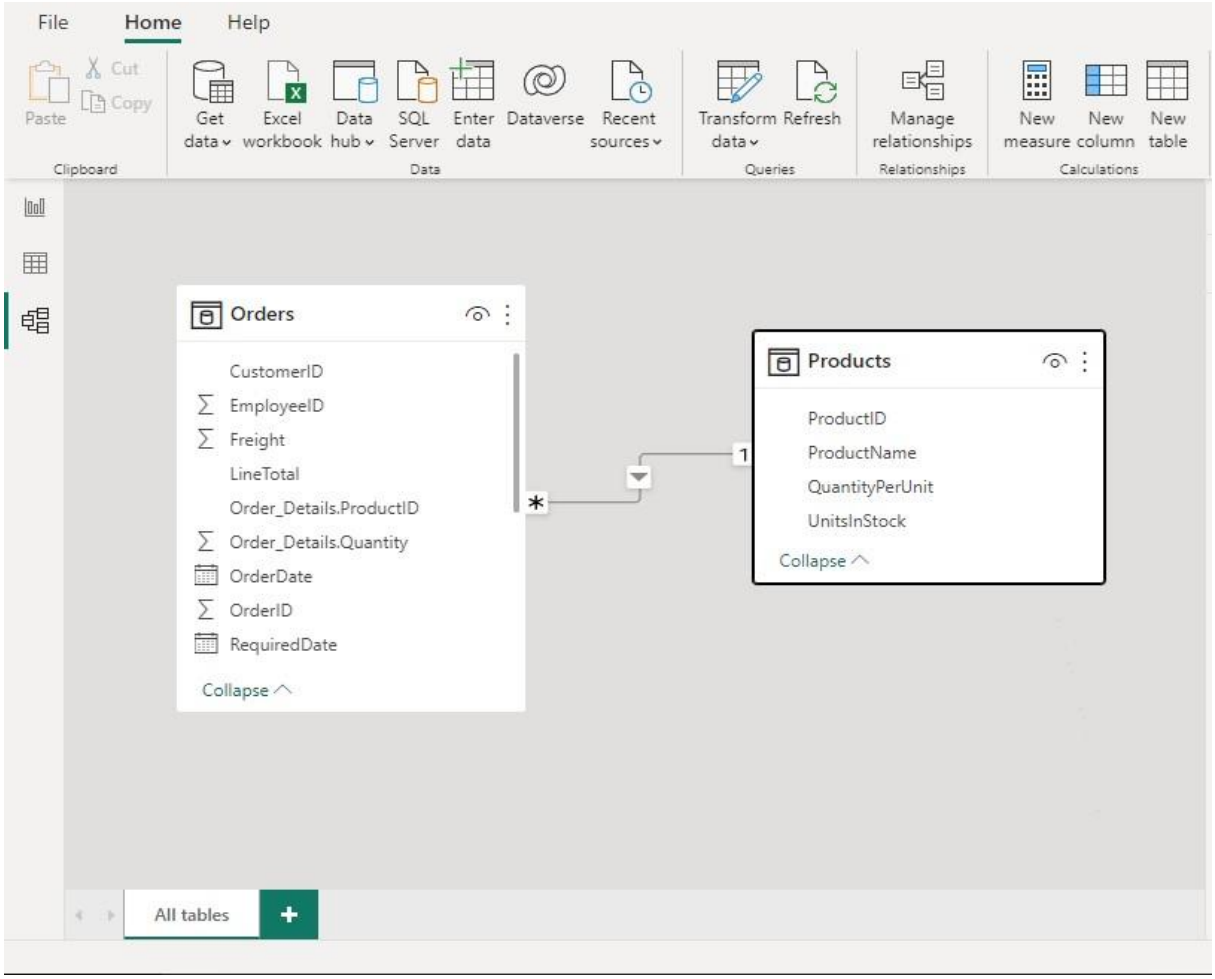
5. When we attempt to create the relationship, we see that one already exists! As shown in the Create Relationship dialog (by the shaded columns), the ProductsID fields in each query already have an established relationship.



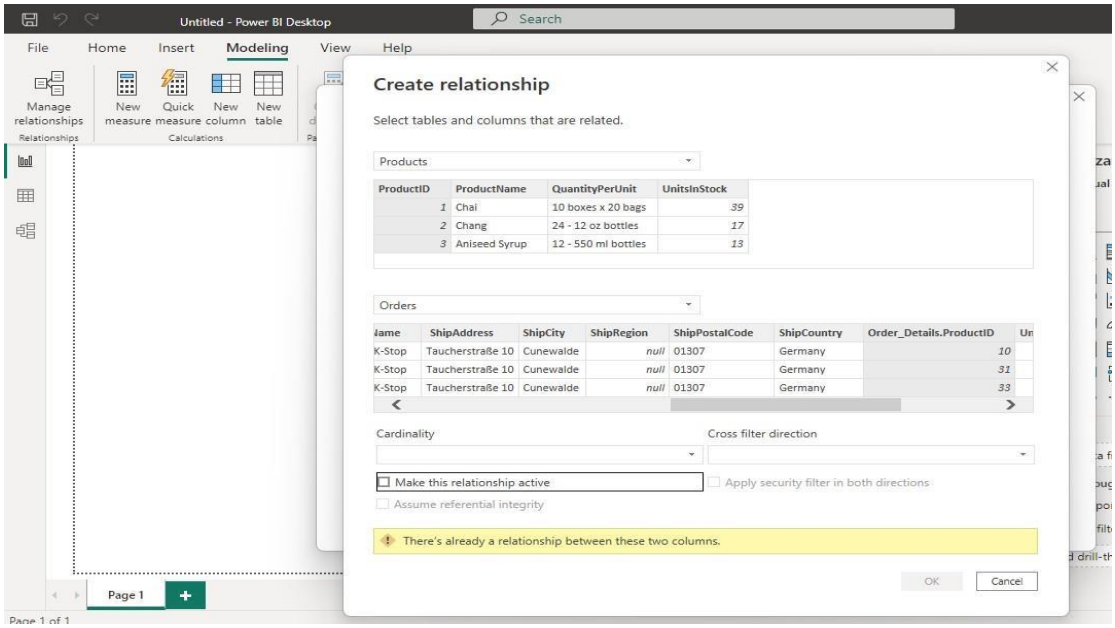
6. Select Cancel, and then select Relationship view in Power BI Desktop.



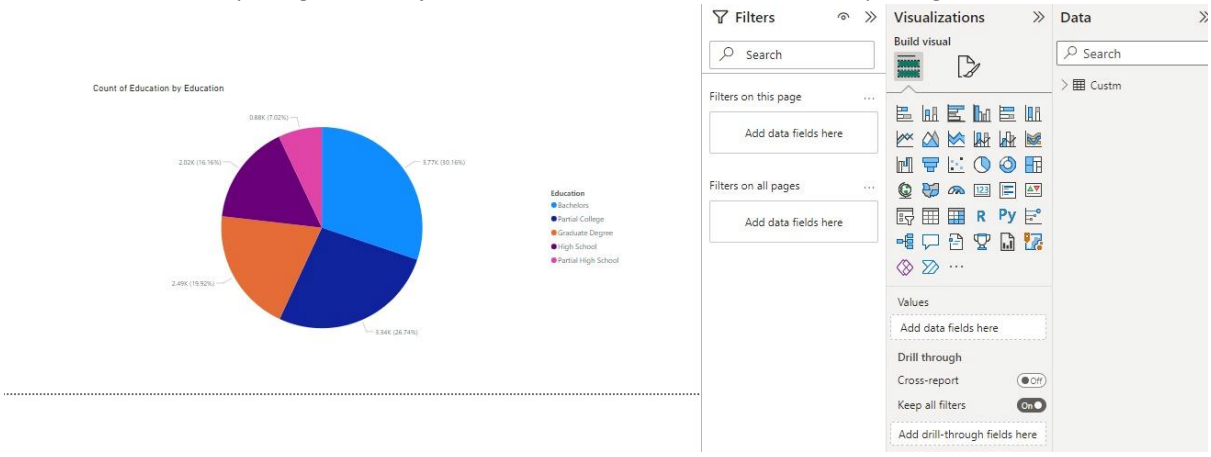
7. We see the following, which visualizes the relationship between the queries.



8. When you double-click the arrow on the line that connects the to queries, an Edit Relationship dialog appears.

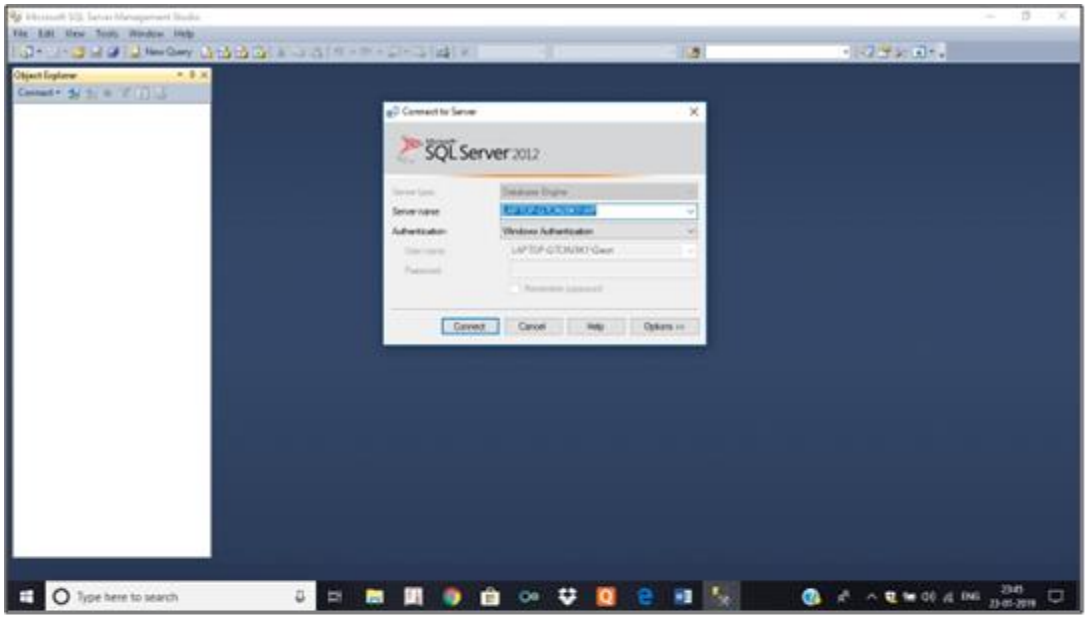


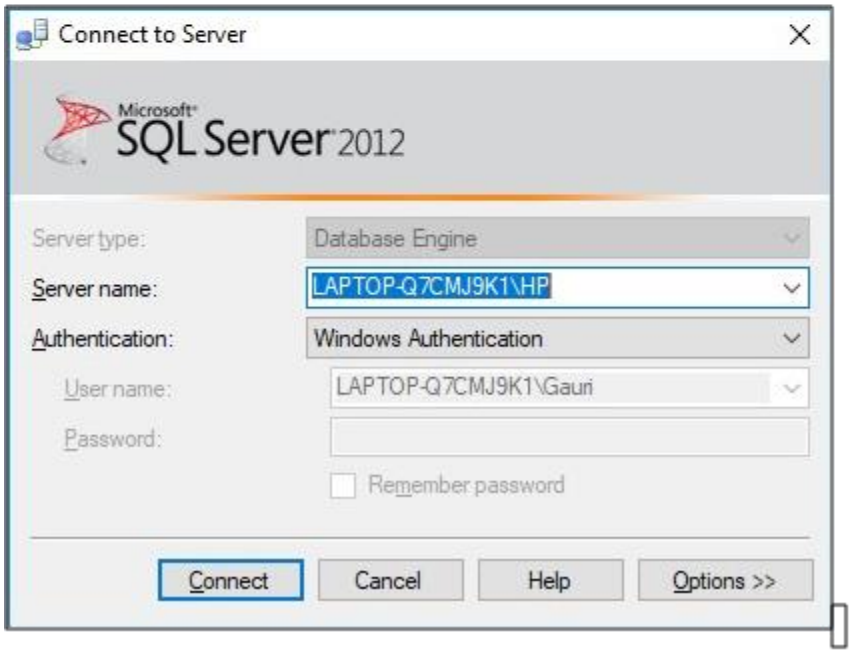
9. No need to make any changes, so we'll just select Cancel to close the Edit Relationship dialog



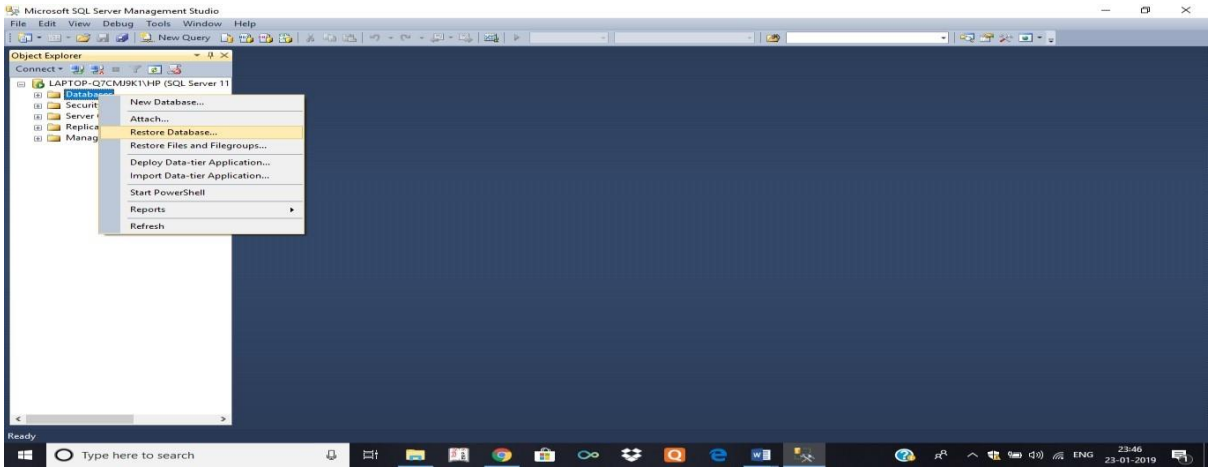
PRACTICAL 6

AIM: Perform the Extraction Transformation and Loading (ETL) process to construct the database in the SQL server.

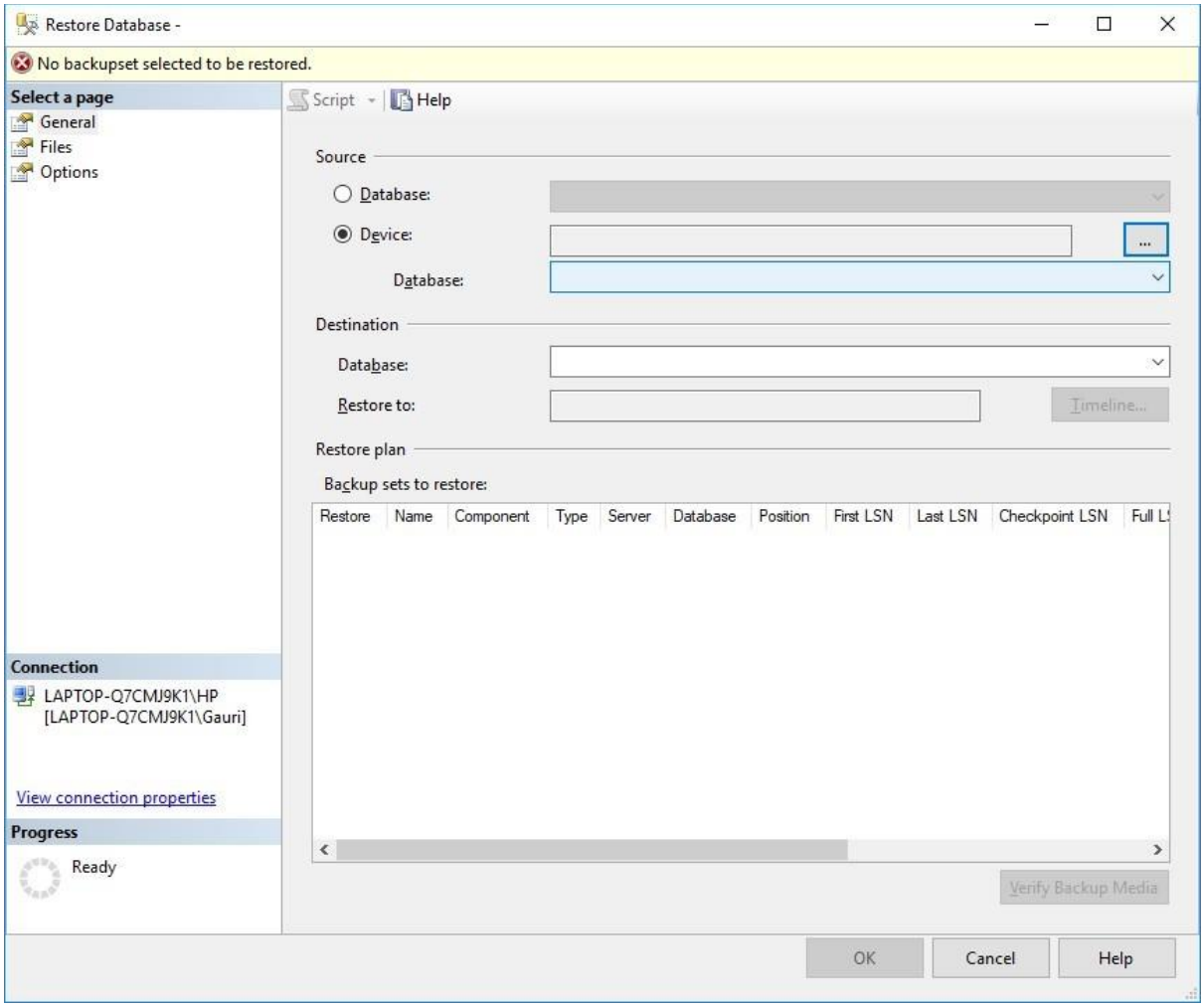


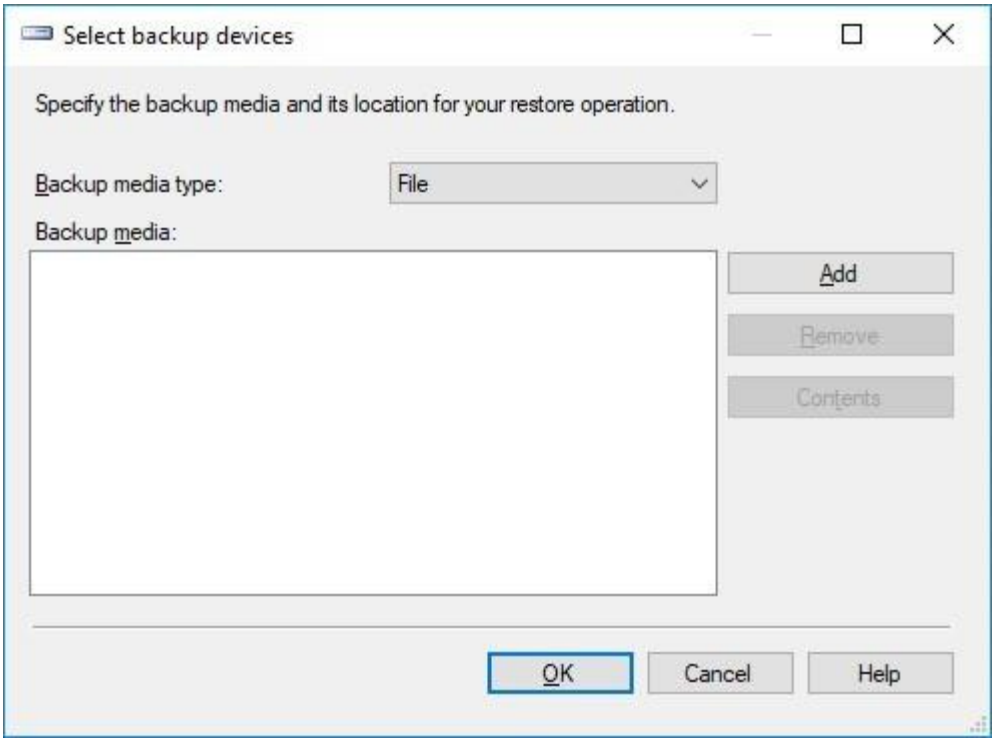


Step 2: Right click on Databases → Restore Database

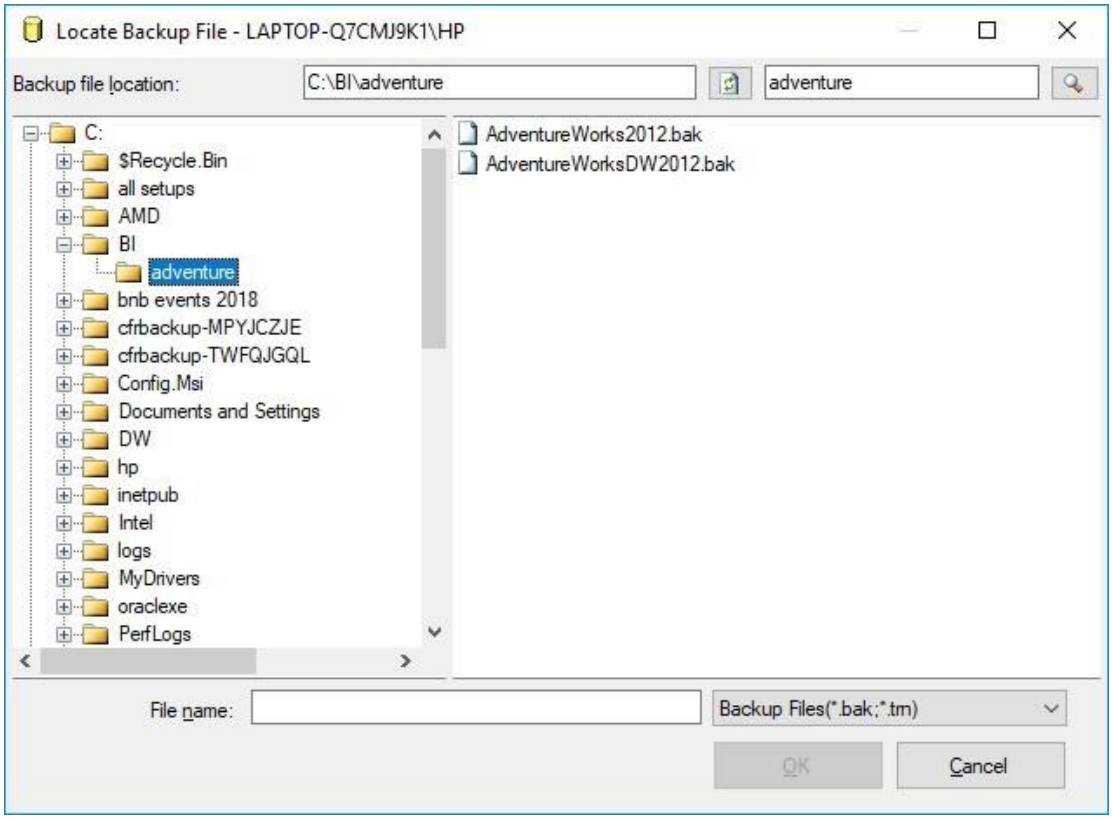


Step 3: Select Device → click on icon towards end of device box.

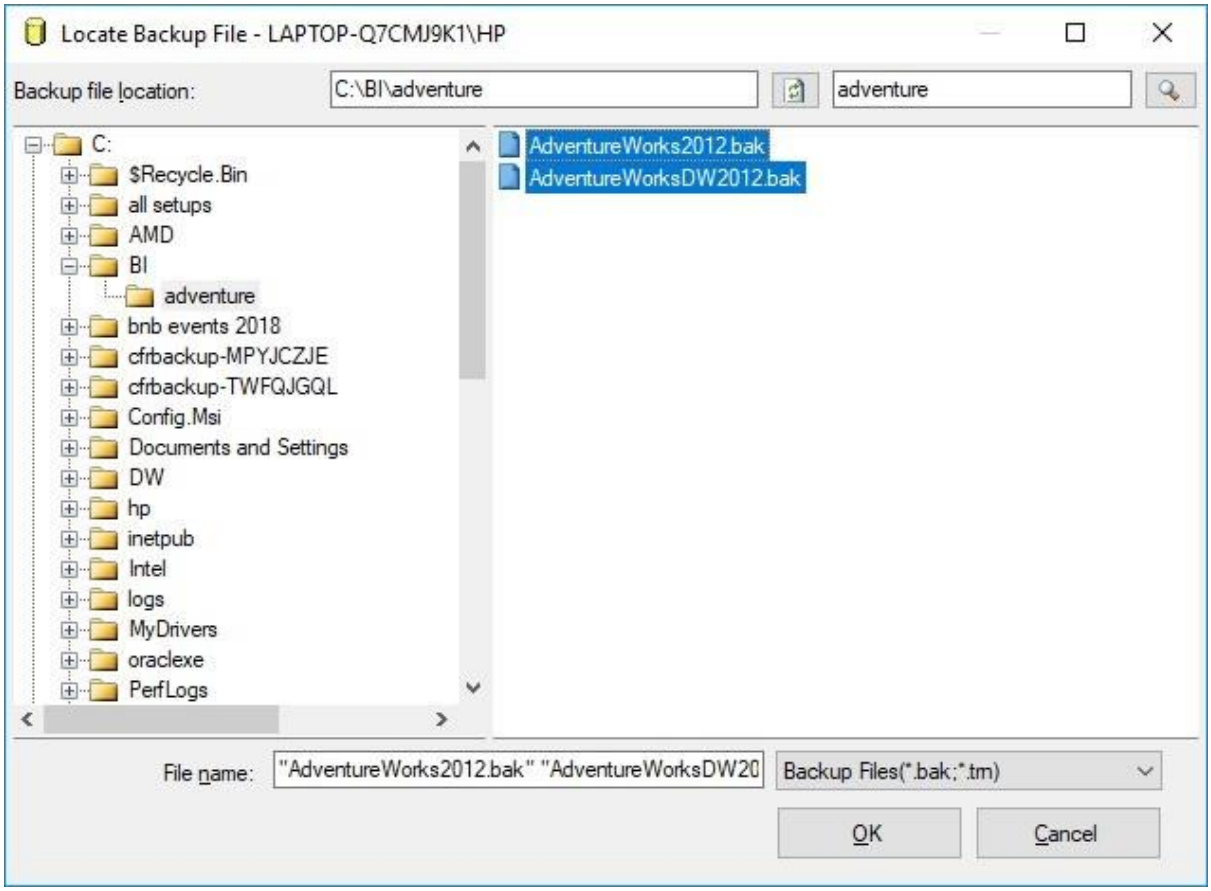




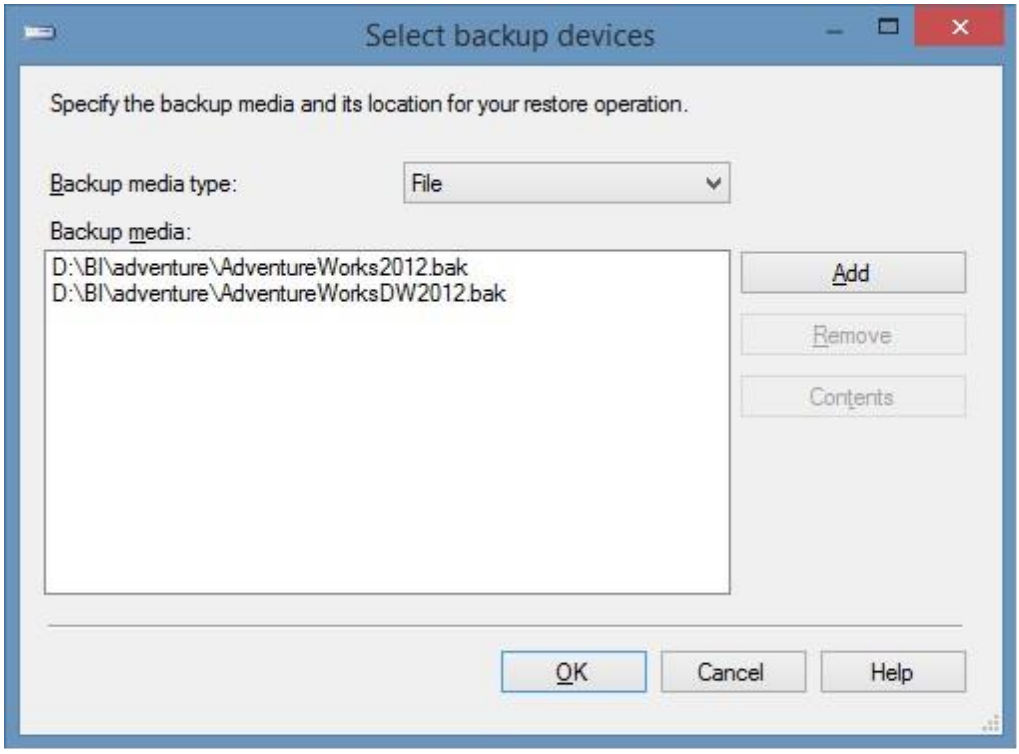
Step 4: Click on Add → Select path of backup files

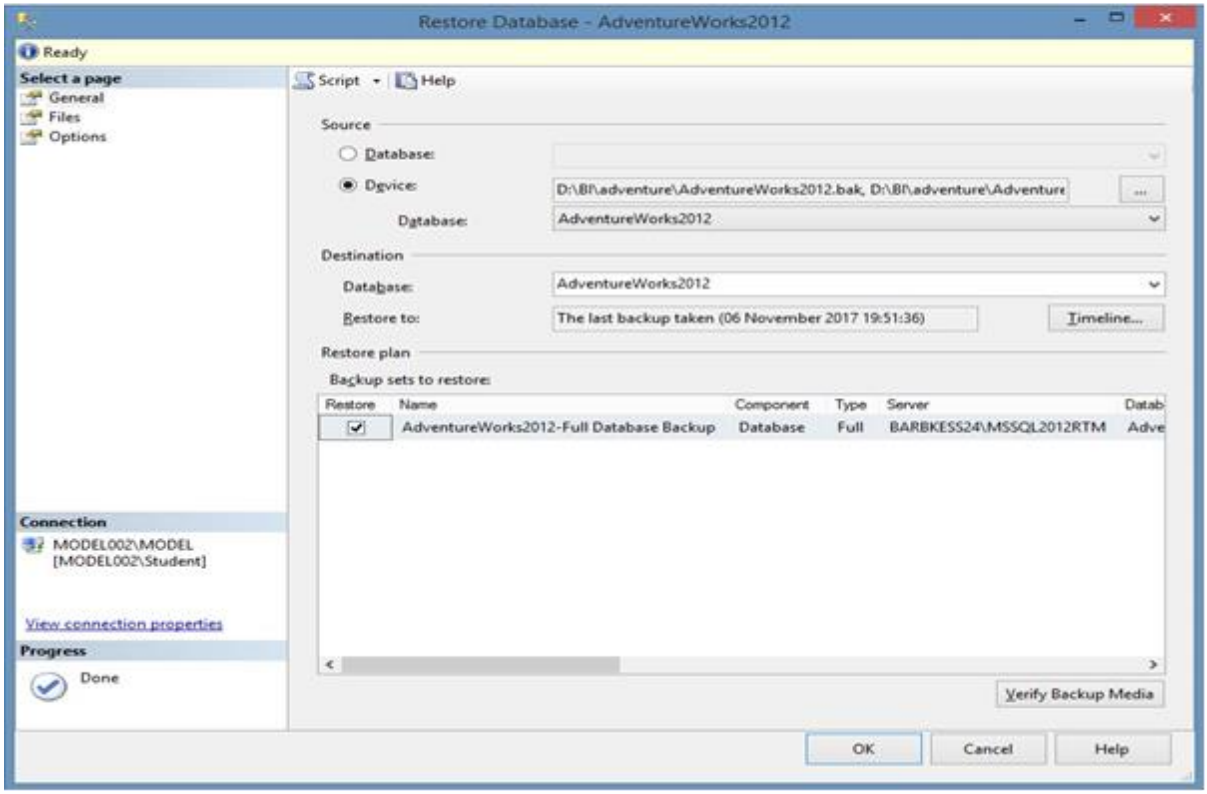


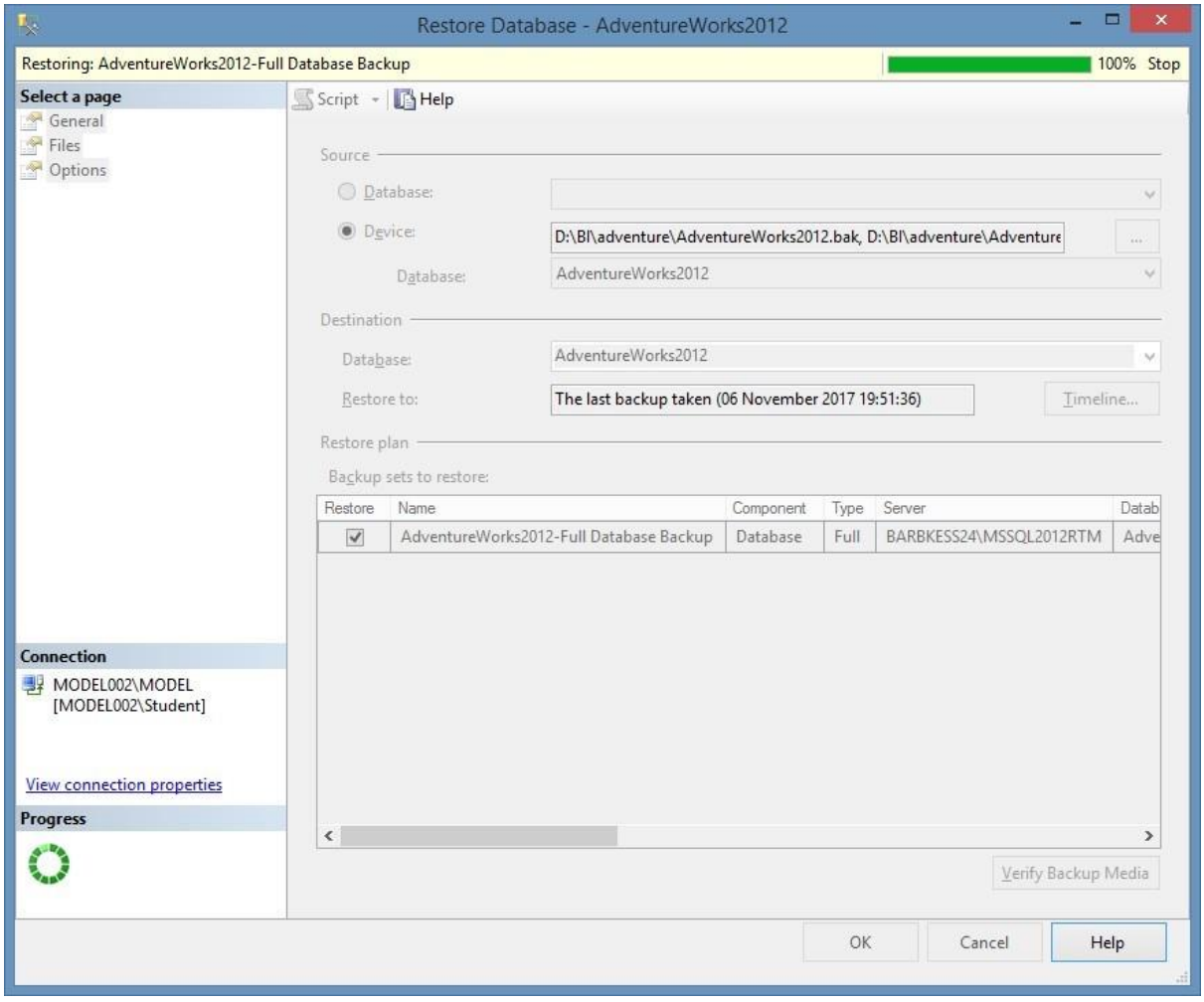
Step 5: Select both files at a time

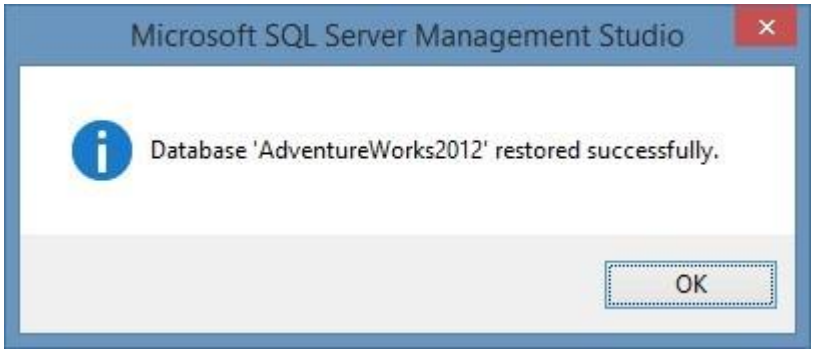


Step 6: Click ok and in select backup devices window Add both files of AdventureWorks

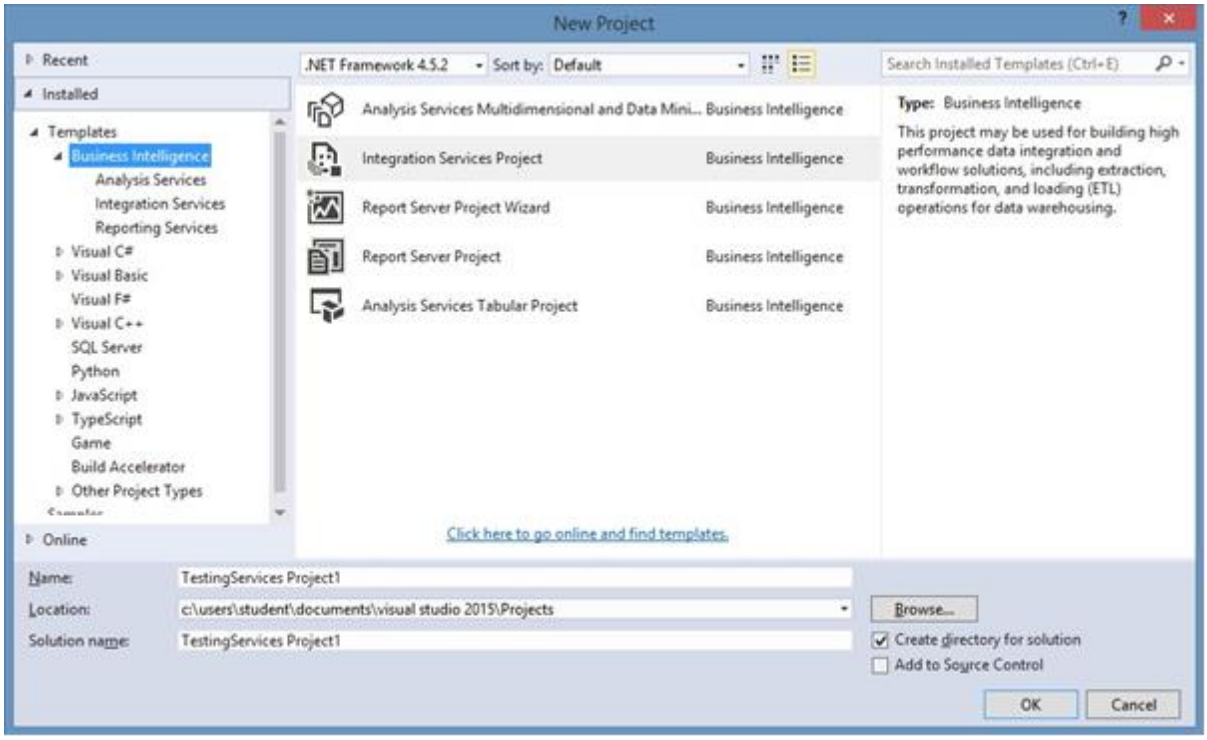

Select backup devices
Specify the backup media and its location for your restore operation.
Backup media type:
File
Backup media:
D:\BI\adventure\AdventureWorks2012.bak
D:\BI\adventure\AdventureWorksDW2012.bak
Add
Remove
Contents
OK
Cancel
Help

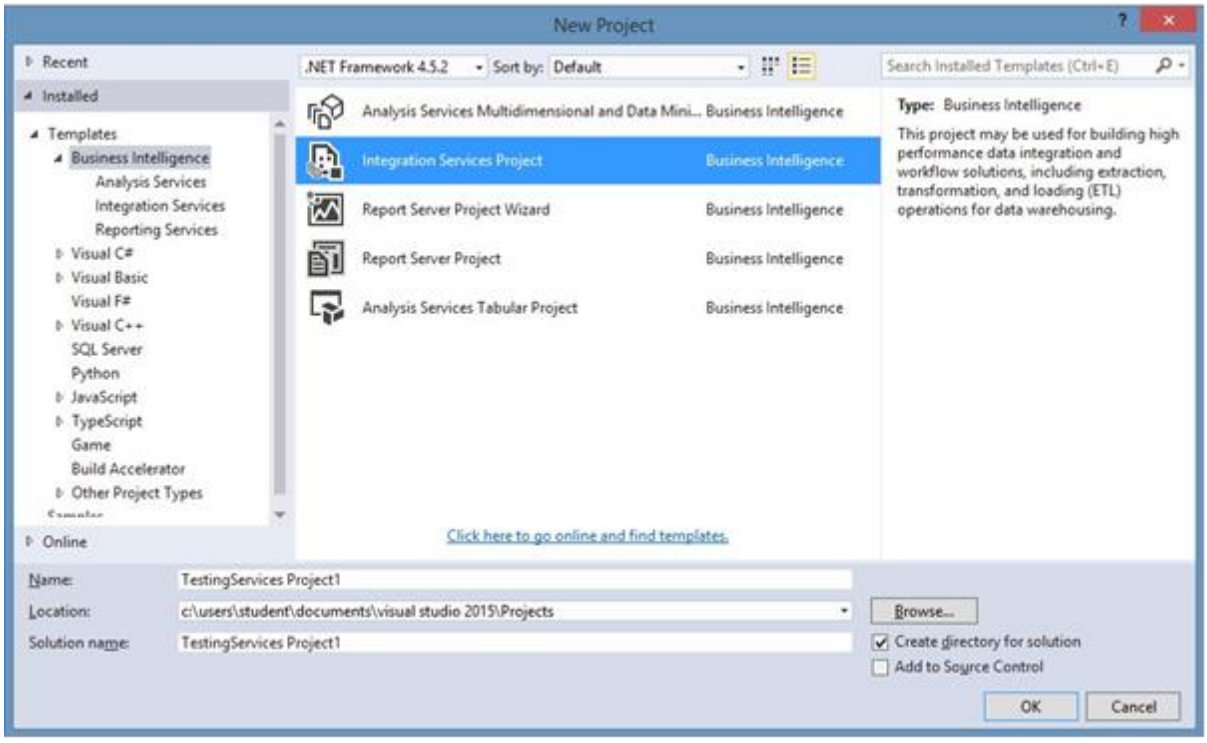

Restore Database - AdventureWorks2012
Ready
Select a page
General
Files
Options
Script
Help
Source
Database:
Device:
D:\BI\adventure\AdventureWorks2012.bak, D:\BI\adventure\Adventure
Database:
AdventureWorks2012
Destination
Database:
AdventureWorks2012
Restore to:
The last backup taken (06 November 2017 19:51:36)
Timeline...
Restore plan
Backup sets to restore:
Restore
Name
Component
Type
Server
Datab
AdventureWorks2012-Full Database Backup
Database
Full
BARBKESS24\MSSQL2012RTM
Adve
Connection
MODEL002\MODEL
[MODEL002\Student]
View connection properties
Progress
Done
Verify Backup Media
OK
Cancel
Help

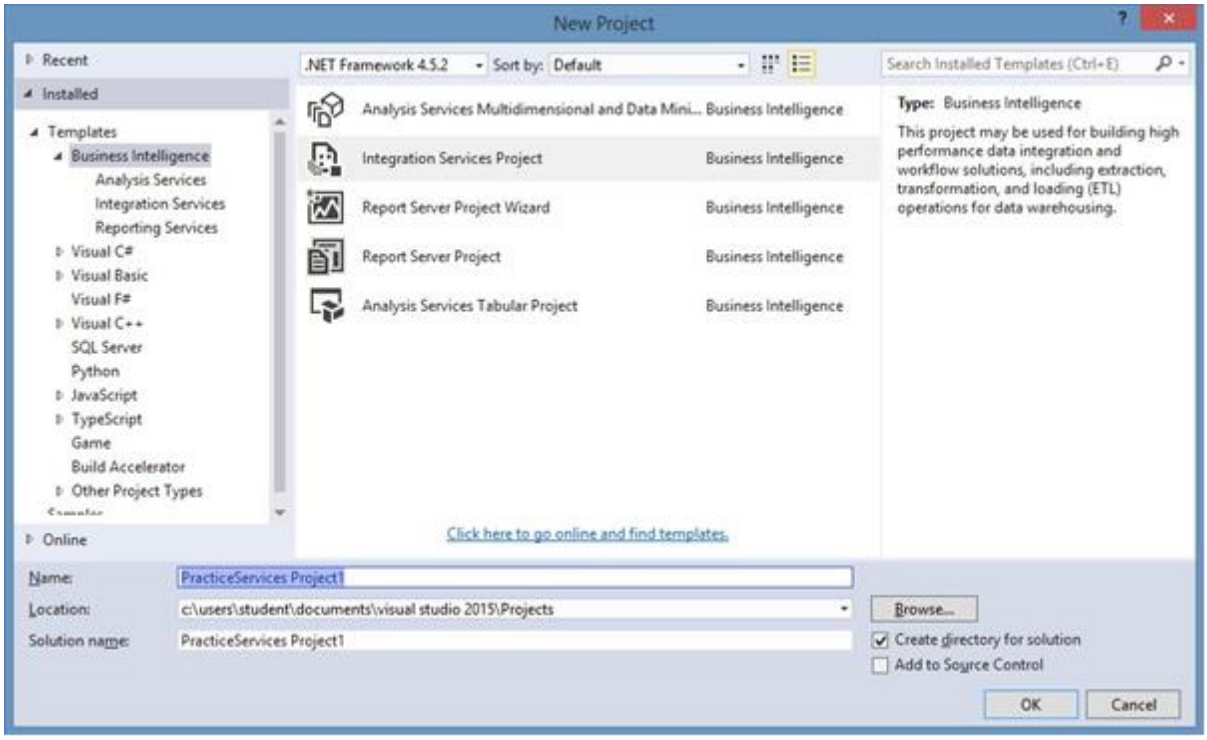


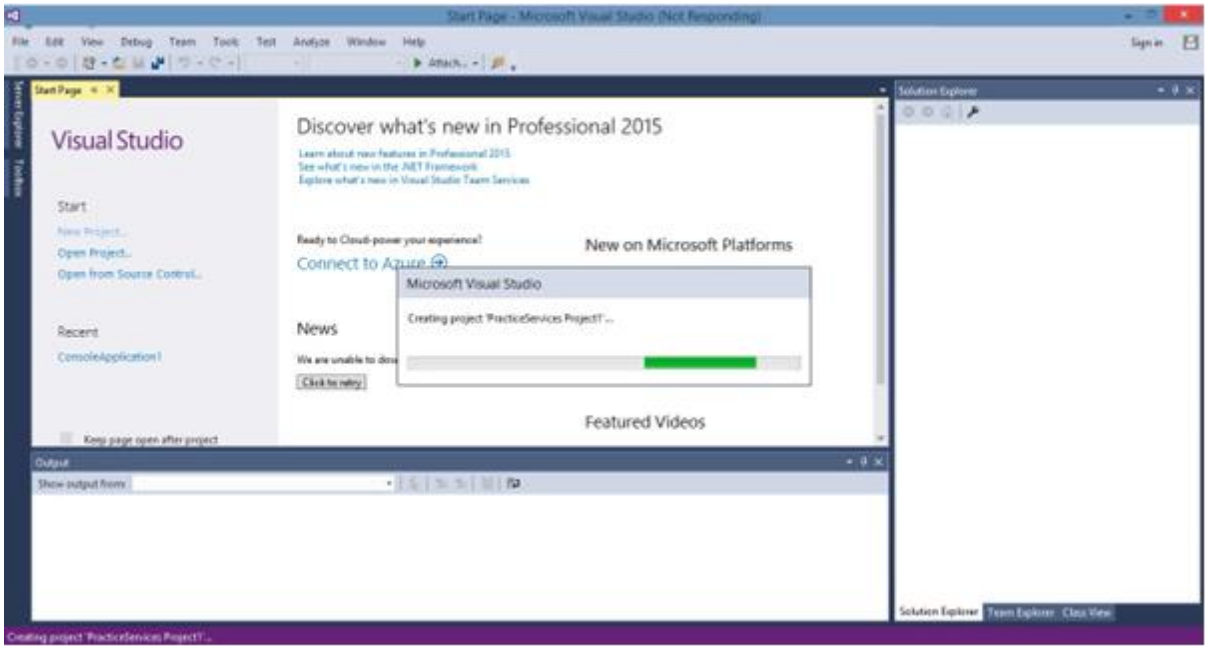


Step 7: Open SQL Server Data Tools
Select File → New → Project → Business Intelligence → Integration Services Project & give appropriate project name.

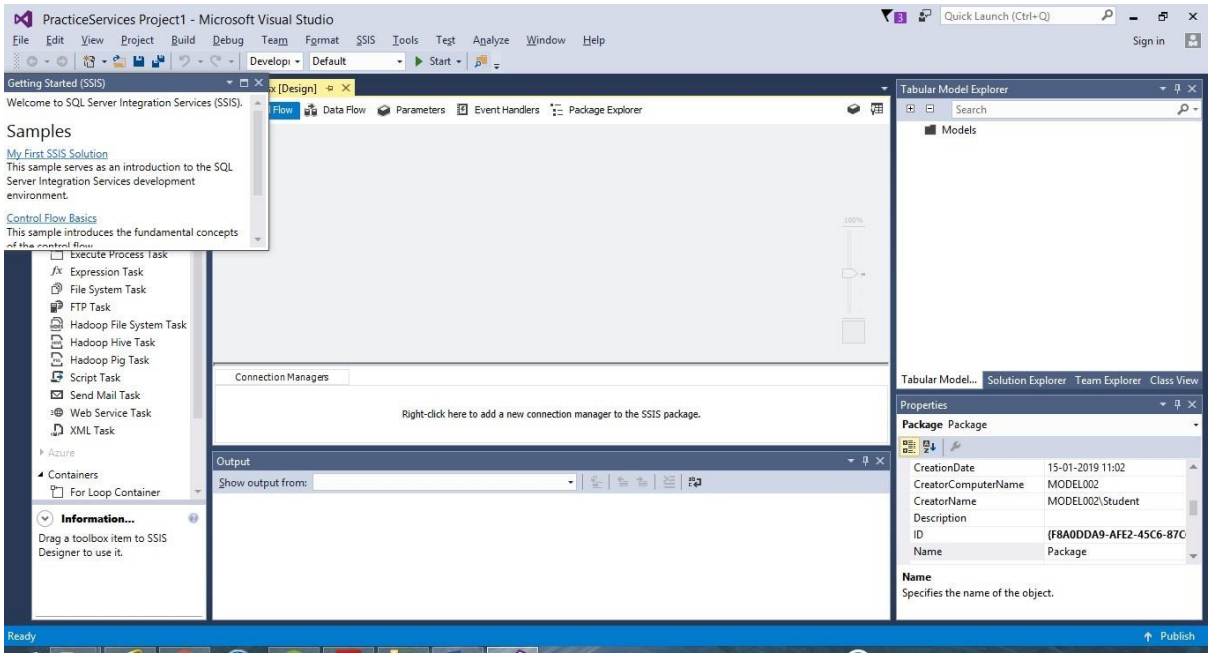

New Project
Recent
Installed
Templates
Business Intelligence
Analysis Services
Integration Services
Reporting Services
Visual C#
Visual Basic
Visual F#
Visual C++
SQL Server
Python
JavaScript
TypeScript
Game
Build Accelerator
Other Project Types
Online
.NET Framework 4.5.2
Sort by: Default
Search Installed Templates (Ctrl+E)
Analysis Services Multidimensional and Data Mini... Business Intelligence
Integration Services Project Business Intelligence
Report Server Project Wizard Business Intelligence
Report Server Project Business Intelligence
Analysis Services Tabular Project Business Intelligence
Type: Business Intelligence
This project may be used for building high performance data integration and workflow solutions, including extraction, transformation, and loading (ETL) operations for data warehousing.
Click here to go online and find templates.
Name: TestingServices Project1
Location: c:\users\student\documents\visual studio 2015\Projects
Solution name: TestingServices Project1
Browse...
Create directory for solution
Add to Source Control
OK
Cancel


New Project
Recent
Installed
Templates
Business Intelligence
Analysis Services
Integration Services
Reporting Services
Visual C#
Visual Basic
Visual F#
Visual C++
SQL Server
Python
JavaScript
TypeScript
Game
Build Accelerator
Other Project Types
Online
.NET Framework 4.5.2
Sort by: Default
Search Installed Templates (Ctrl+E)
Analysis Services Multidimensional and Data Mini... Business Intelligence
Integration Services Project Business Intelligence
Report Server Project Wizard Business Intelligence
Report Server Project Business Intelligence
Analysis Services Tabular Project Business Intelligence
Type: Business Intelligence
This project may be used for building high performance data integration and workflow solutions, including extraction, transformation, and loading (ETL) operations for data warehousing.
Click here to go online and find templates.
Name: TestingServices Project1
Location: c:\users\student\documents\visual studio 2015\Projects
Solution name: TestingServices Project1
Browse...
Create directory for solution
Add to Source Control
OK
Cancel

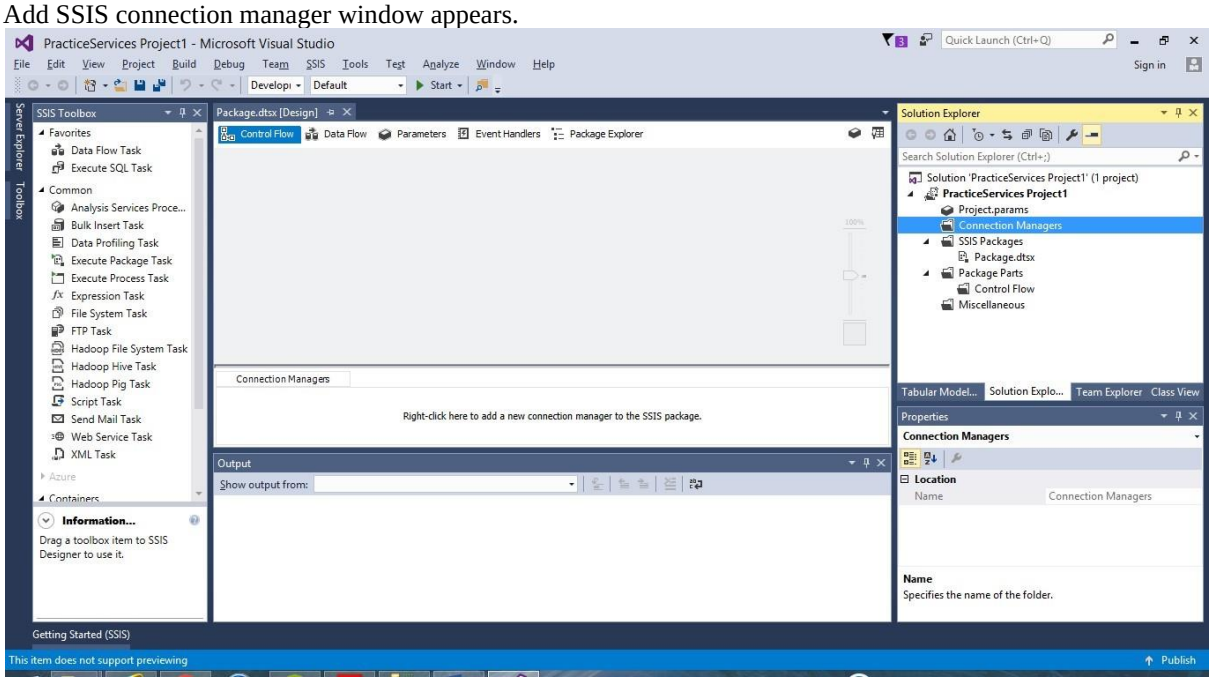




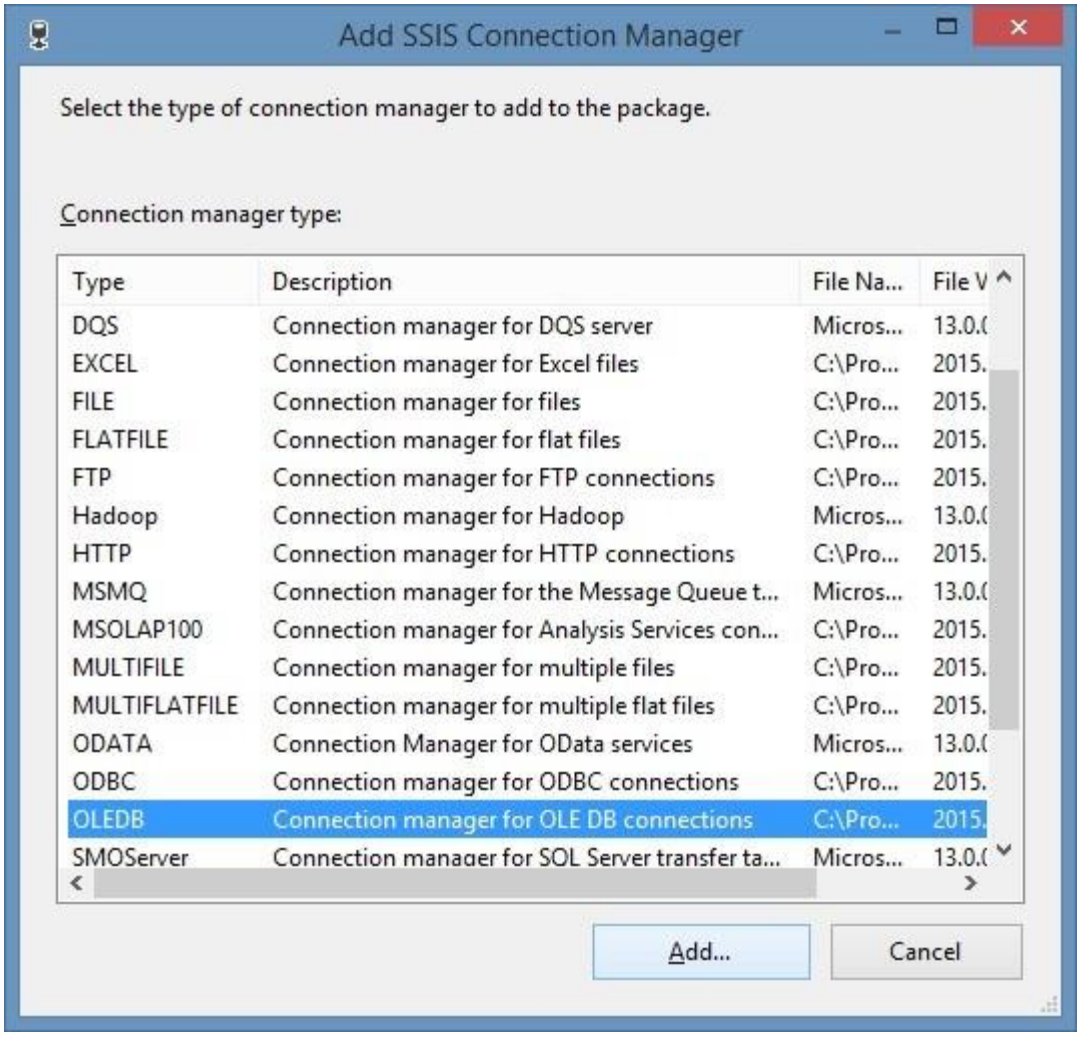
Environment consists of SQL Server Integration Services (SSIS)



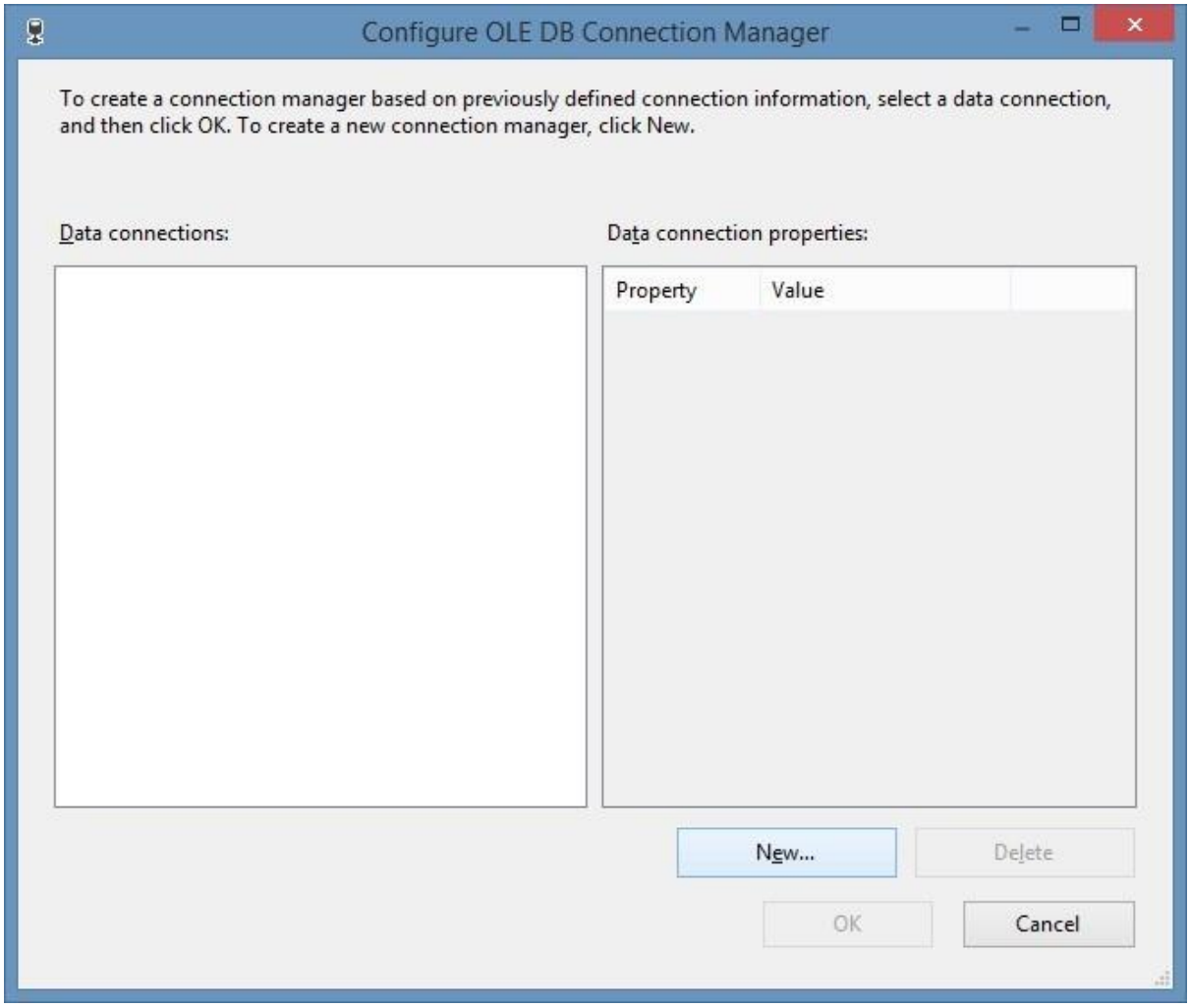
Step 8: Right click on Connection Managers in solution explorer and click on New Connection Manager. Add SSIS connection manager window appears.



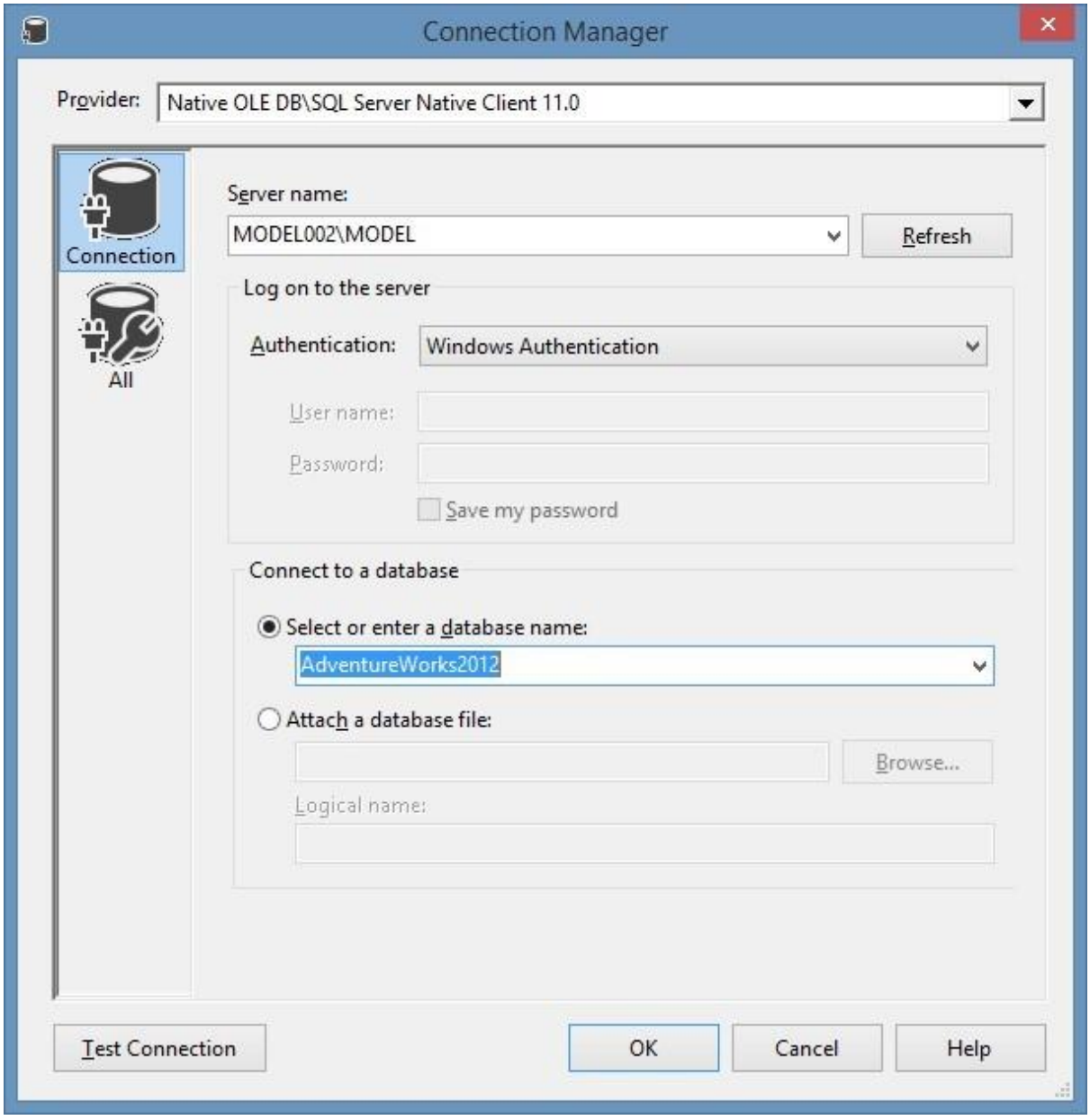
Step 9: Select OLE DB Connection Manager and Click on Add



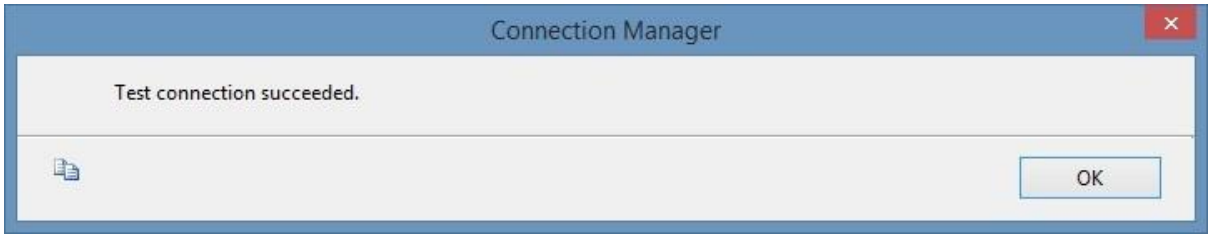
Step 10: Configure OLE DB Connection Manager window appears → Click on New



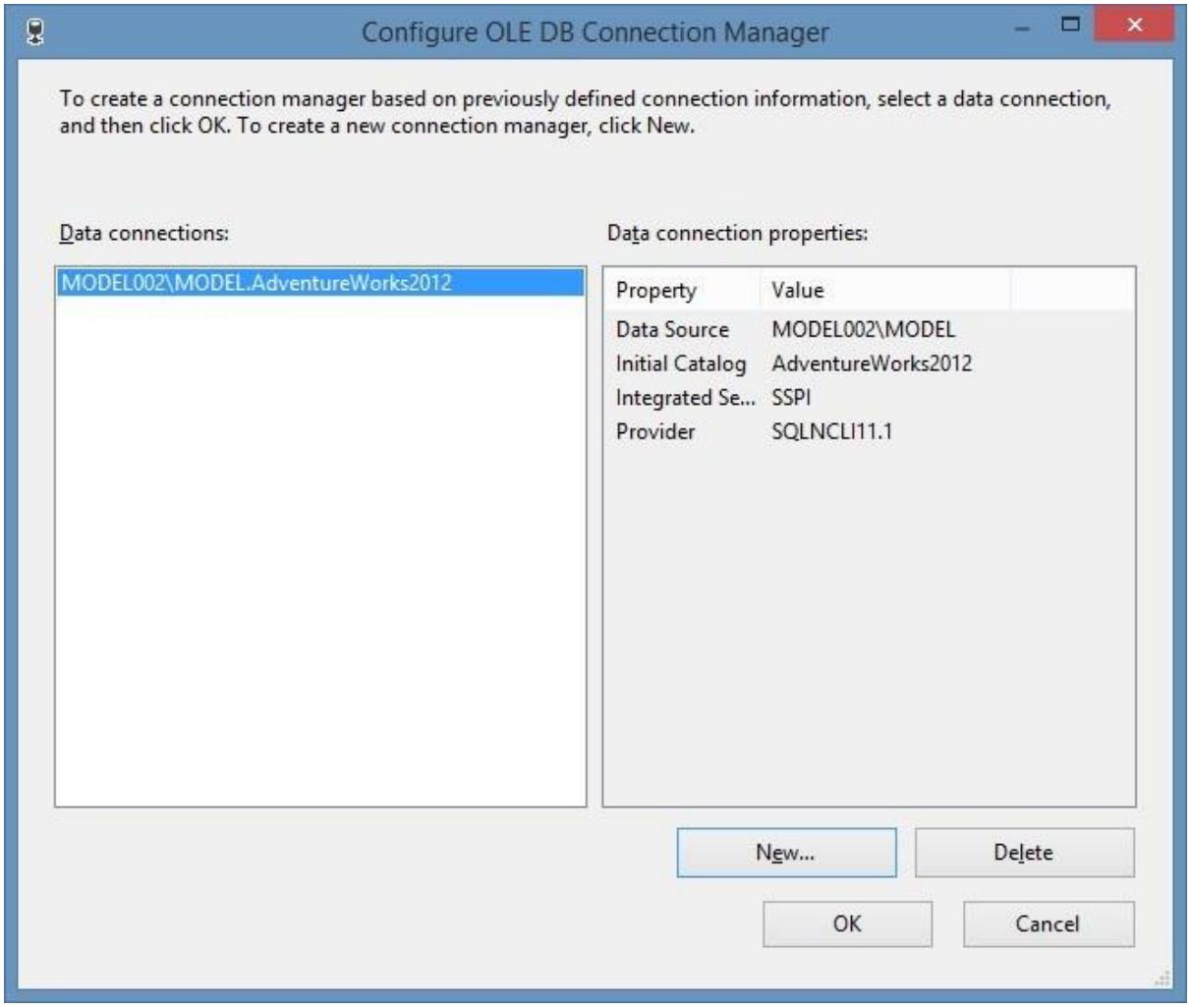
Step 11: Select Server name (as per your machine) from drop down and database name and click on Test connection.



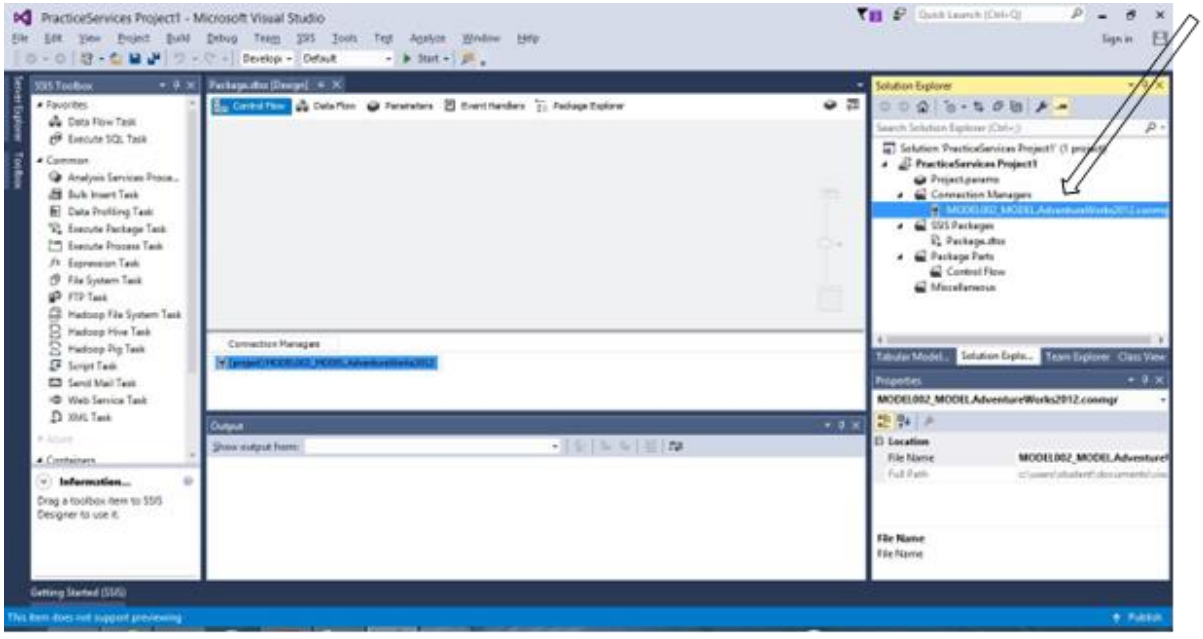
If test connection succeeded click on OK.



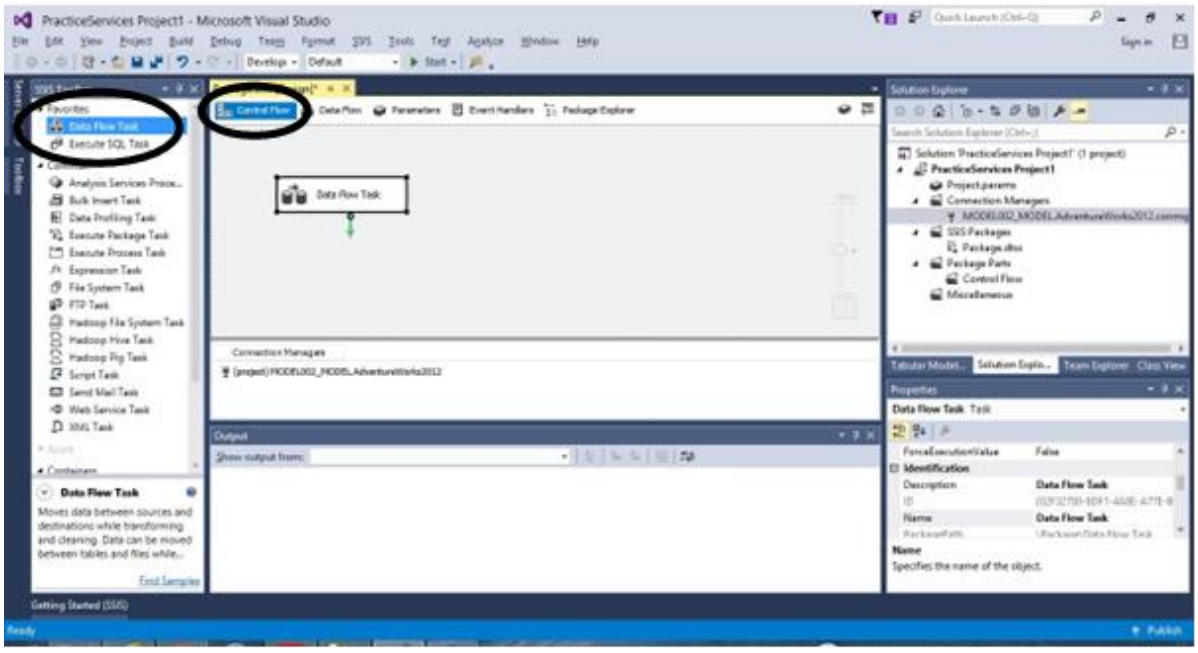
Step 12: Click on OK



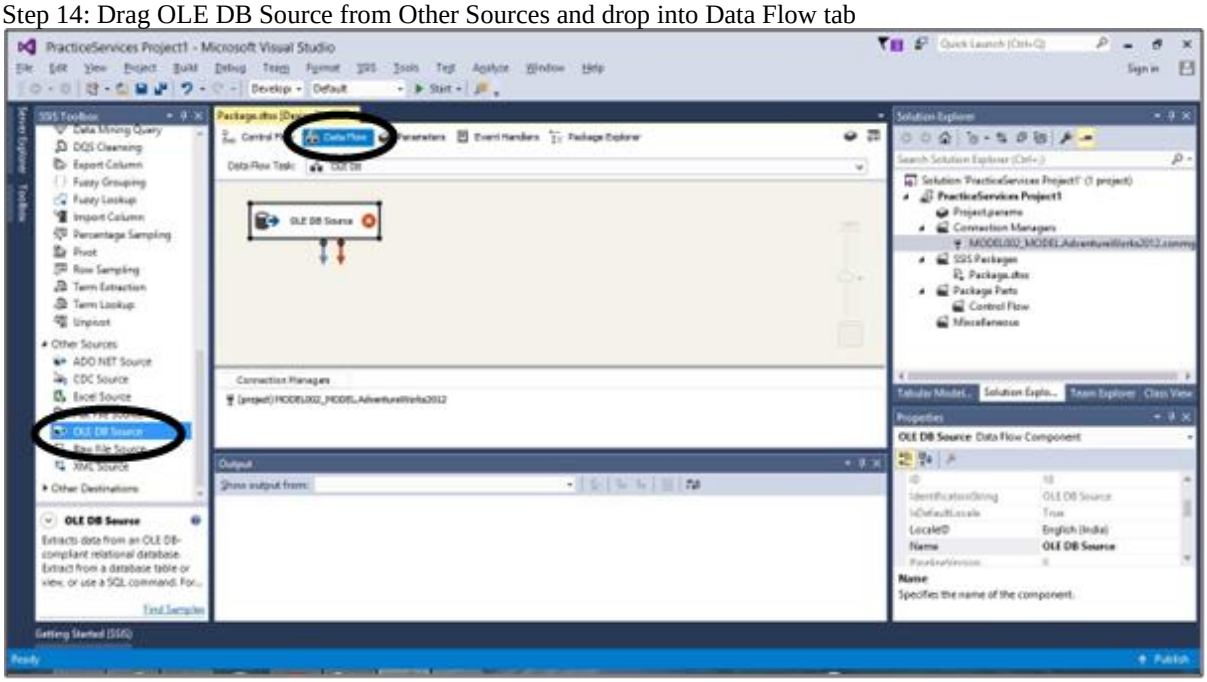
Connection is added to connection manager



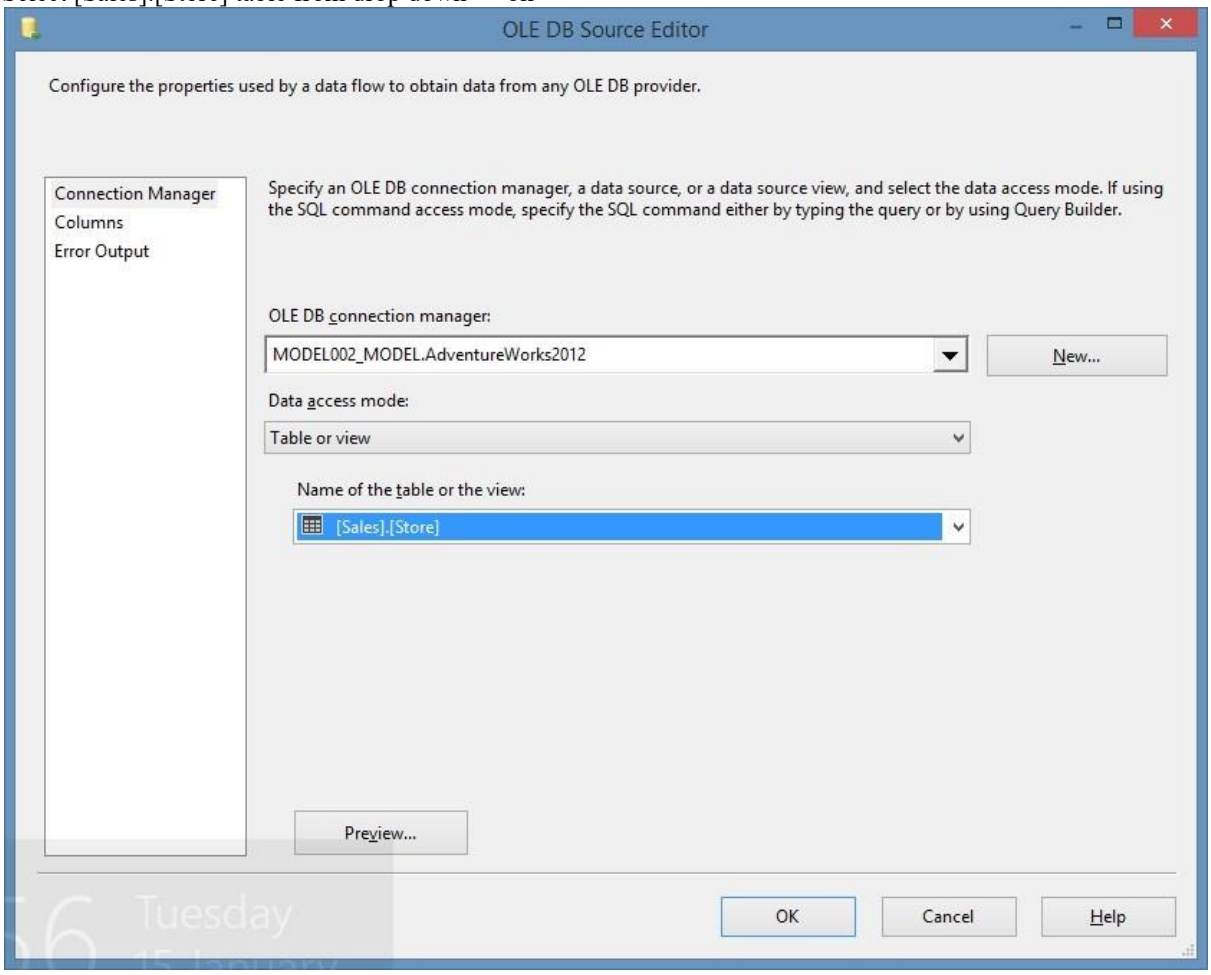
Step 13: Drag and drop Data Flow Task in Control Flow tab



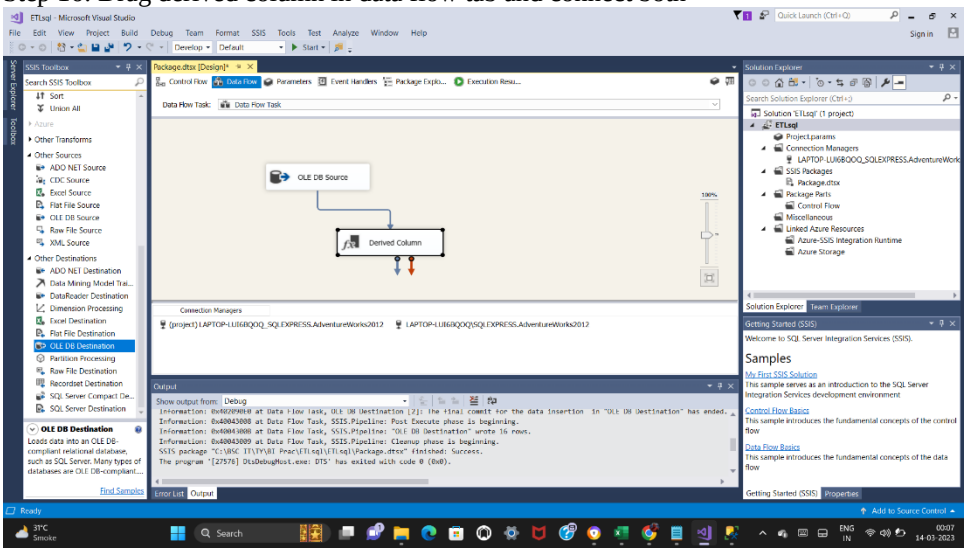
Step 14: Drag OLE DB Source from Other Sources and drop into Data Flow tab



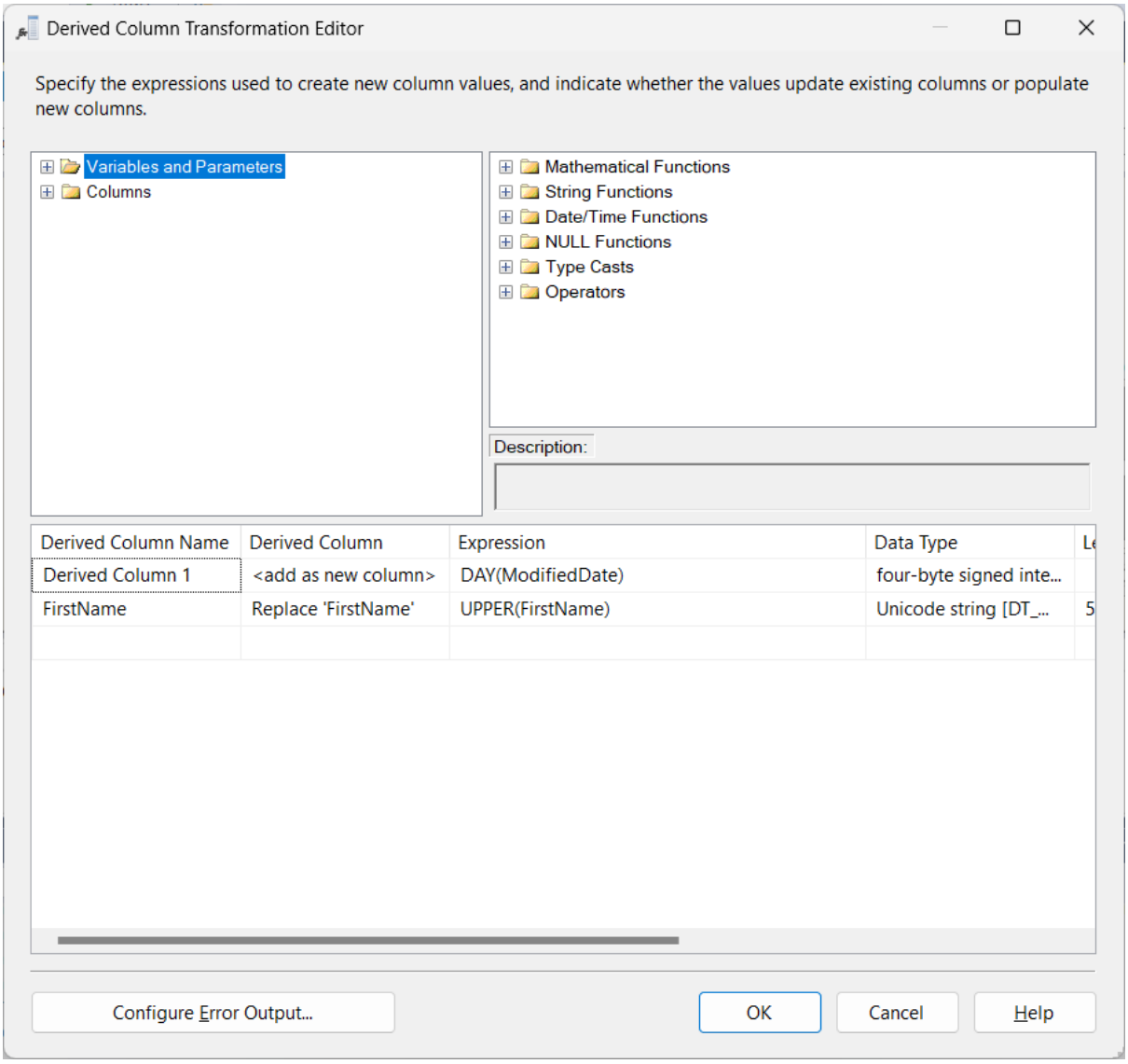
Step 15: Double click on OLE DB source → OLE DB Source Editor appears→ click on New to add connection manager.
Select [Sales].[Store] table from drop down → ok



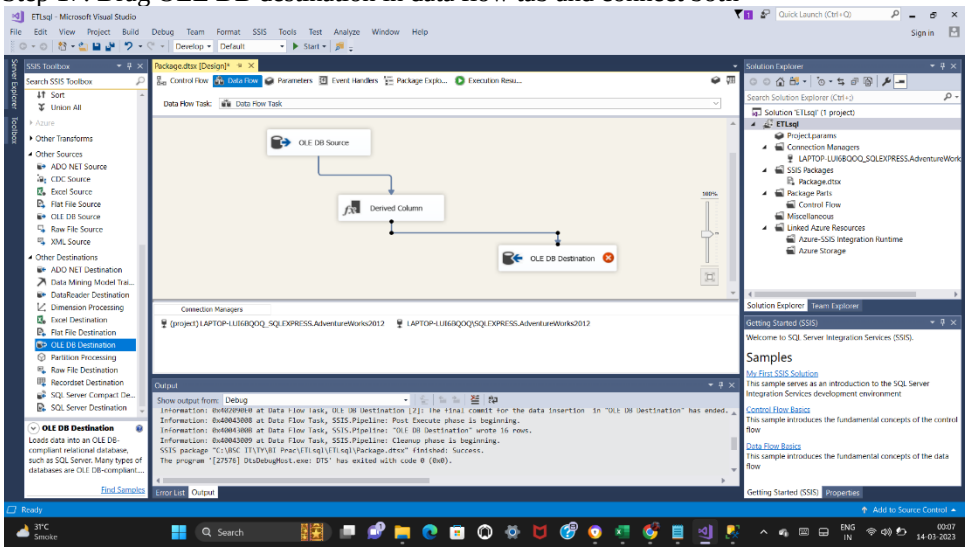
Step 16: Drag derived column in data flow tab and connect both



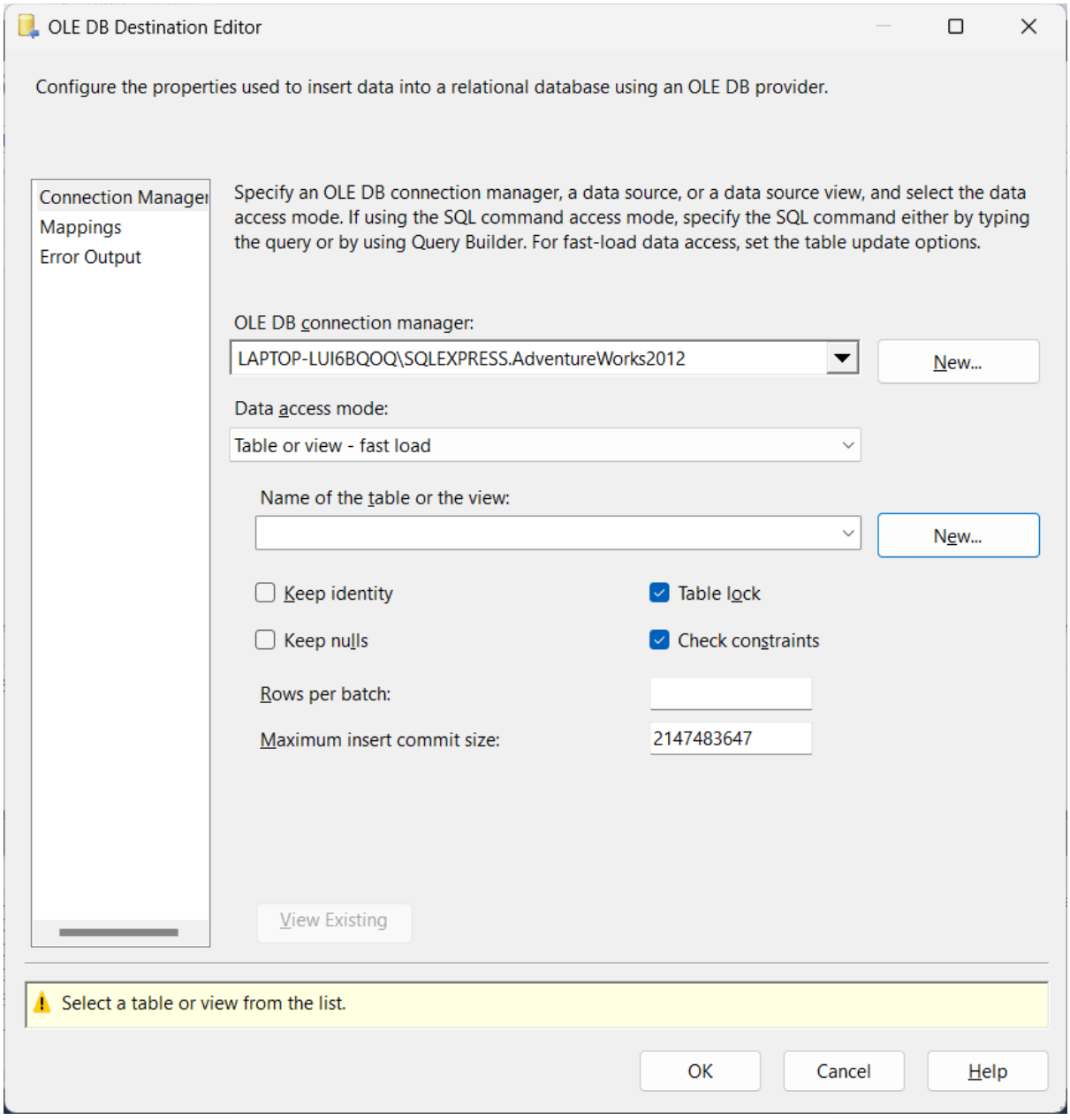
Double click on derived column → ok

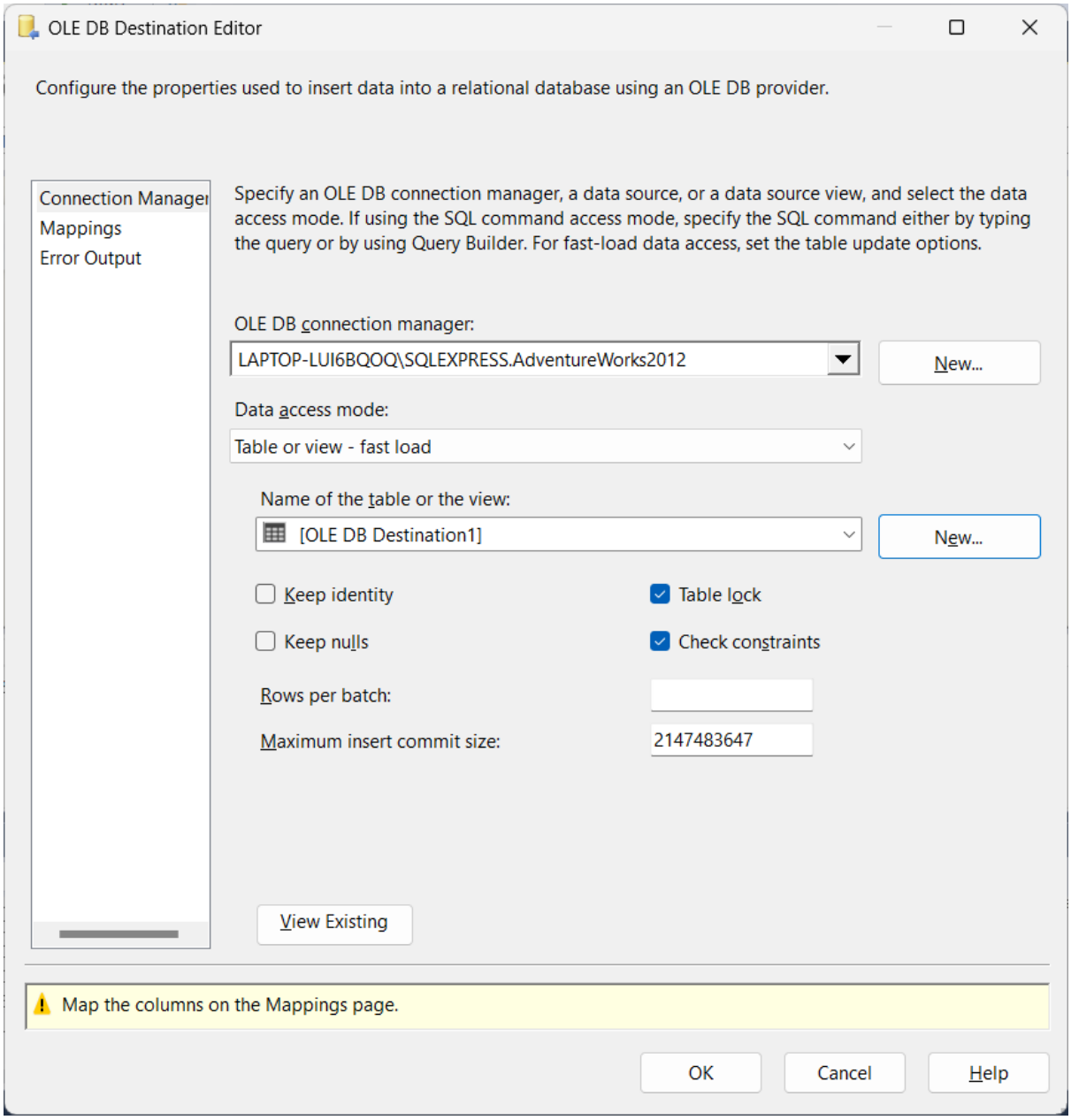


Step 17: Drag OLE DB destination in data flow tab and connect both

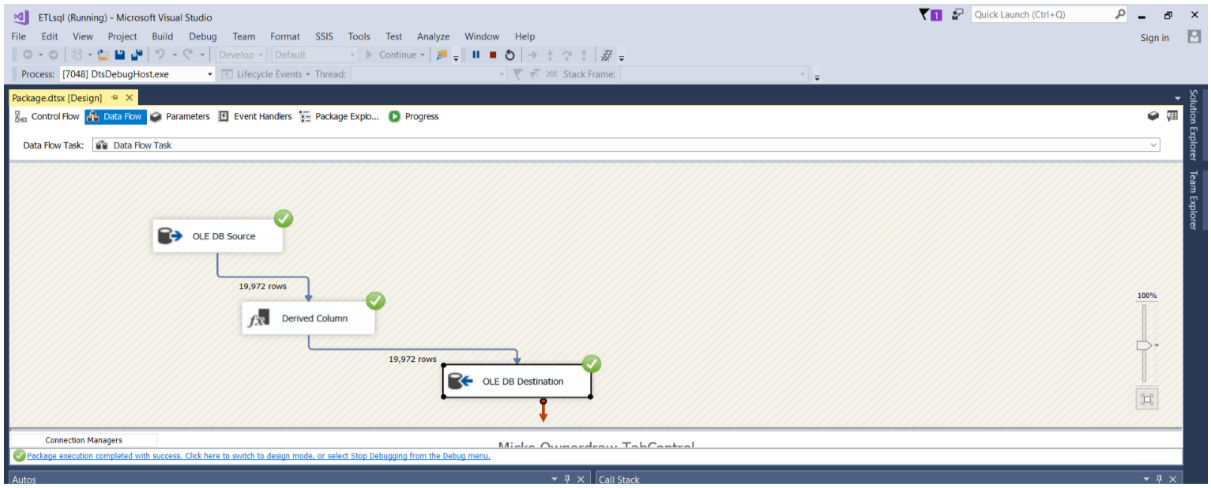


Step 18: Double click on OLE DB destination
Click on New to run the query to get [OLE DB Destination] in Name of the table or the view.

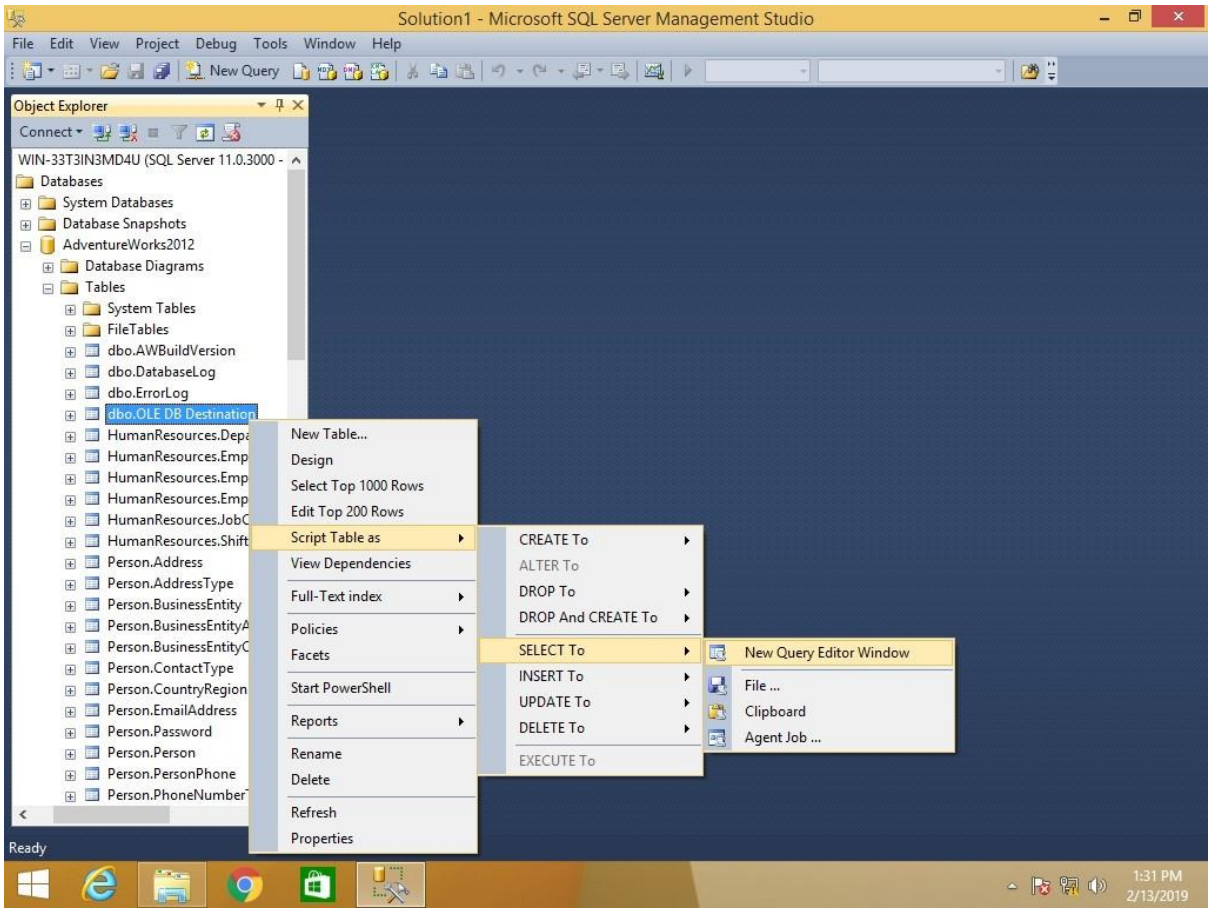




Click on ok
Step 19: Click on start.

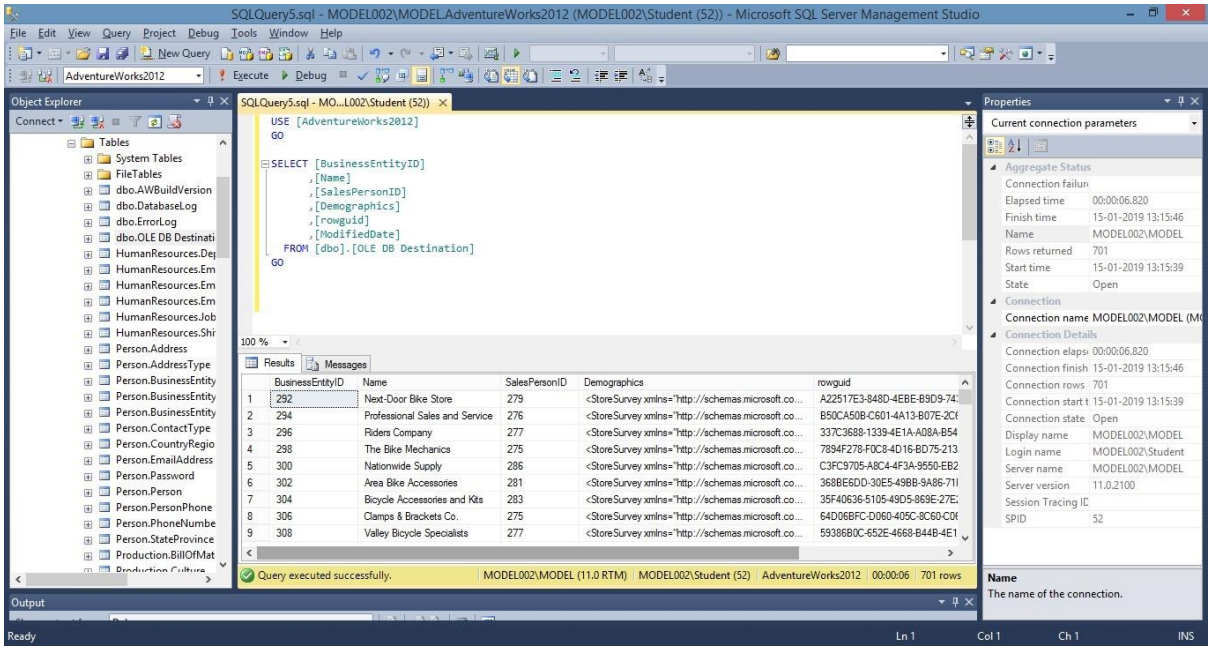


Step 20: Go to SQL Server Management Studio
In database tab → Adventure works → Right click on [dbo].[OLE DB Destination] → Scrip Table as → SELECT To → New Query Editor Window



Step 21: Execute following query to get output.

```
USE [AdventureWorks2012]
GO
        SELECT [BusinessEntityID], [Name], [SalesPersonID], [Demographics], [rowguid],
        [ModifiedDate]   FROM [dbo]. [OLE DB Destination]
GO
```


SQLQuery5.sql - MODEL002\MODEL.AdventureWorks2012 (MODEL002\Student (52)) - Microsoft SQL Server Management Studio
File Edit View Query Project Debug Tools Window Help
New Query
AdventureWorks2012
Execute Debug
Object Explorer
Connect
Tables
System Tables
FileTables
dbo.AWBuildVersion
dbo.DatabaseLog
dbo.ErrorLog
dbo.OLE DB Destinati
HumanResources.De
HumanResources.Em
HumanResources.Em
HumanResources.Em
HumanResources.Job
HumanResources.Shi
Person.Address
Person.AddressType
Person.BusinessEntity
Person.BusinessEntity
Person.BusinessEntity
Person.ContactType
Person.CountryRegio
Person.EmailAddress
Person.Password
Person.Person
Person.PersonPhone
Person.PhoneNumbe
Person.StateProvince
Production.BillOfMat
SQLQuery5.sql - MO...L002\Student (52))
USE [AdventureWorks2012]
GO
SELECT [BusinessEntityID]
,[Name]
,[SalesPersonID]
,[Demographics]
,[rowguid]
,[ModifiedDate]
FROM [dbo].[OLE DB Destination]
GO
100 %
Results
Messages
BusinessEntityID Name SalesPersonID Demographics rowguid
1 292 Next-Door Bike Store 279 <StoreSurvey xmlns="http://schemas.microsoft.co... A22517E3-848D-4EBE-B9D9-74
2 294 Professional Sales and Service 276 <StoreSurvey xmlns="http://schemas.microsoft.co... B50CA50B-C601-4A13-B07E-2C
3 296 Riders Company 277 <StoreSurvey xmlns="http://schemas.microsoft.co... 337C3688-1339-4E1A-A08A-B54
4 298 The Bike Mechanics 275 <StoreSurvey xmlns="http://schemas.microsoft.co... 7894F278-F0C8-4D16-BD75-213
5 300 Nationwide Supply 286 <StoreSurvey xmlns="http://schemas.microsoft.co... C3FC9705-A8C4-4F3A-9550-EB2
6 302 Area Bike Accessories 281 <StoreSurvey xmlns="http://schemas.microsoft.co... 368BE6DD-30E5-49BB-9A86-71
7 304 Bicycle Accessories and Kits 283 <StoreSurvey xmlns="http://schemas.microsoft.co... 35F40636-5105-49D5-869E-27E
8 306 Clamps & Brackets Co. 275 <StoreSurvey xmlns="http://schemas.microsoft.co... 64D06BFC-D060-405C-8C60-C0
9 308 Valley Bicycle Specialists 277 <StoreSurvey xmlns="http://schemas.microsoft.co... 59386B0C-652E-4668-B44B-4E1
Query executed successfully.
MODEL002\MODEL (11.0 RTM)
MODEL002\Student (52)
AdventureWorks2012
00:00:06
701 rows
Properties
Current connection parameters
Aggregate Status
Connection failur
Elapsed time 00:00:06.820
Finish time 15-01-2019 13:15:46
Name MODEL002\MODEL
Rows returned 701
Start time 15-01-2019 13:15:39
State Open
Connection
Connection name MODEL002\MODEL (M
Connection Details
Connection elaps 00:00:06.820
Connection finish 15-01-2019 13:15:46
Connection rows 701
Connection start t 15-01-2019 13:15:39
Connection state Open
Display name MODEL002\MODEL
Login name MODEL002\Student
Server name MODEL002\MODEL
Server version 11.0.2100
Session Tracing ID
SPID 52
Name
The name of the connection.
Output
Ready
Ln 1
Col 1
Ch 1
INS

PRACTICAL 7

AIM: Create the cube with suitable dimension and fact tables based on

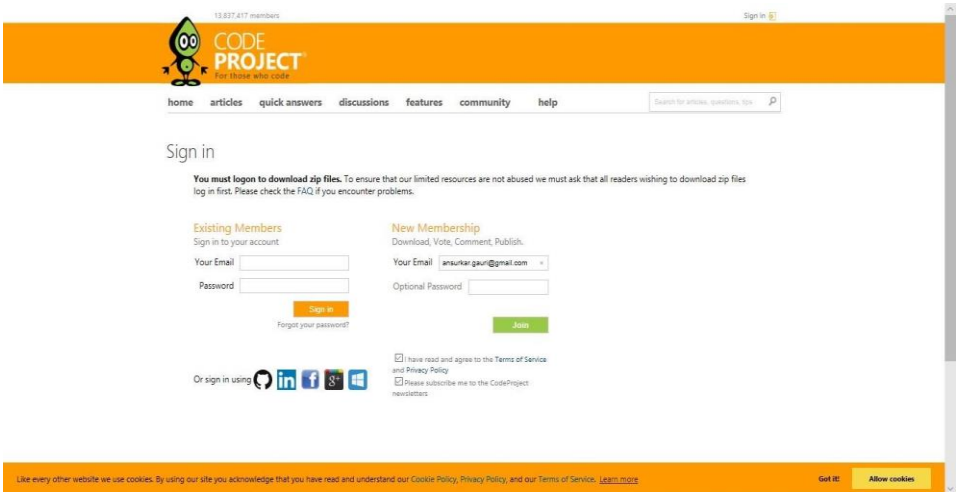
OLAP

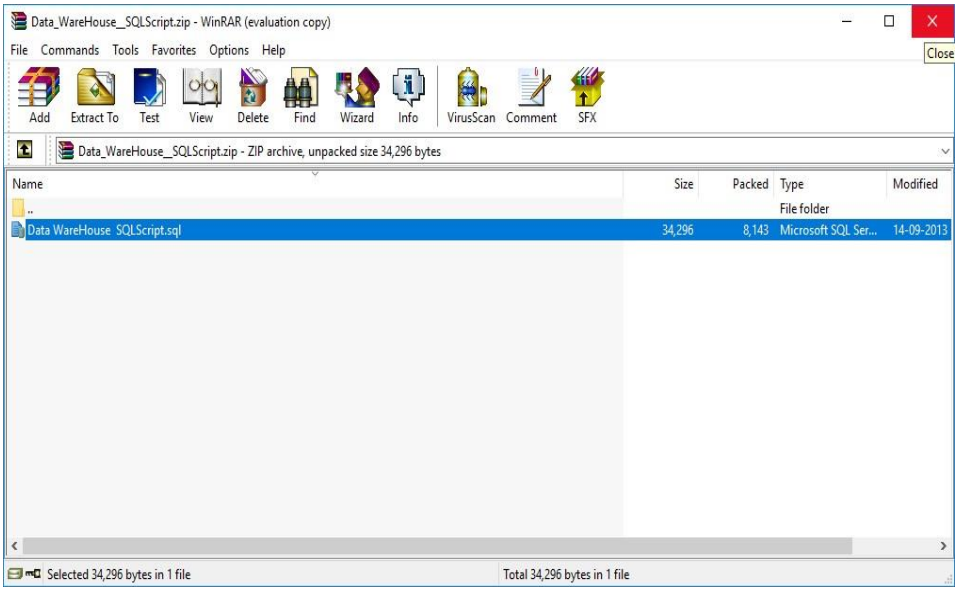
Step 1: Creating Data Warehouse

Let us execute our T-SQL Script to create data warehouse with fact tables, dimensions and populate them with appropriate test values.

Download T-SQL script attached with this article for creation of Sales Data Warehouse or download from this article "Create First Data Warehouse" and run it in your SQL Server.

Downloading "Data_WareHouse_____SQLScript.zip" from the article https://www.codeproject.com/Articles/652108/Create-First-Data-WareHou se





After downloading extract file in folder.

Follow the given steps to run the query in SSMS (SQL Server Management Studio).

1. **Open SQL Server Management Studio 2012**

2. **Connect Database Engine**

Password for sa : admin123 (as given during installation)

Click Connect.

3. **Open New Query editor**

4. **Copy paste Scripts given below in various steps in new query editor window one by one**

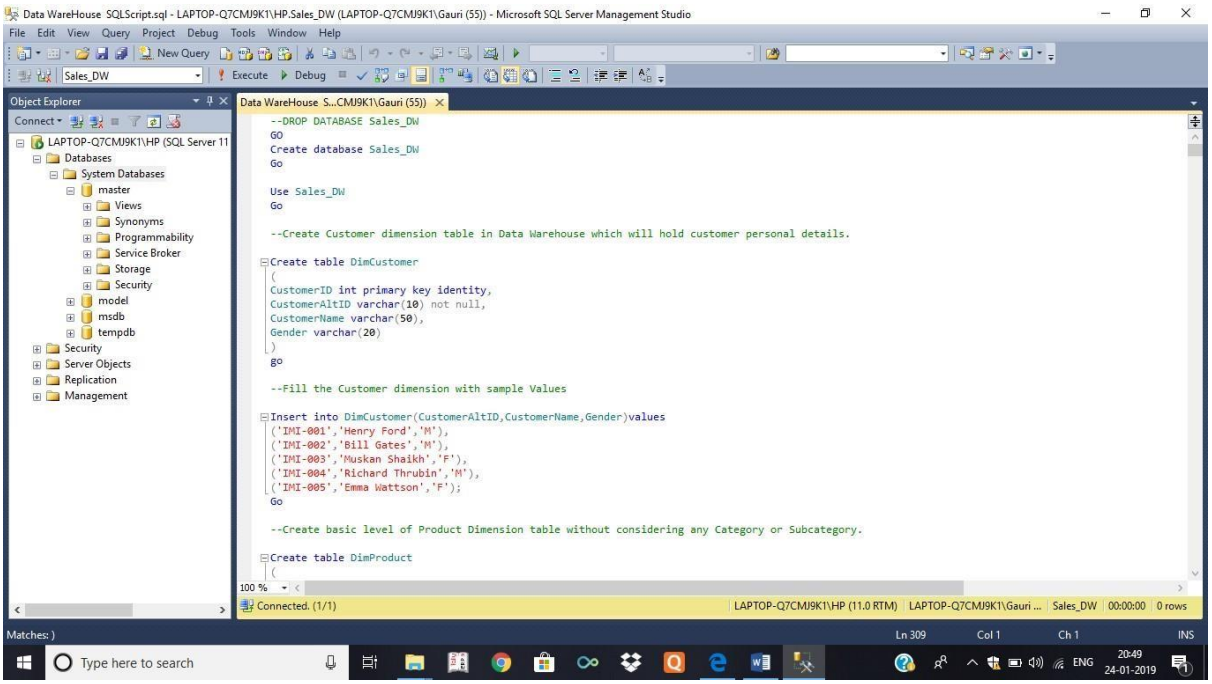
5. **To run the given SQL Script, press F5**

6. **It will create and populate "Sales_DW" database on your SQL Server**

OR

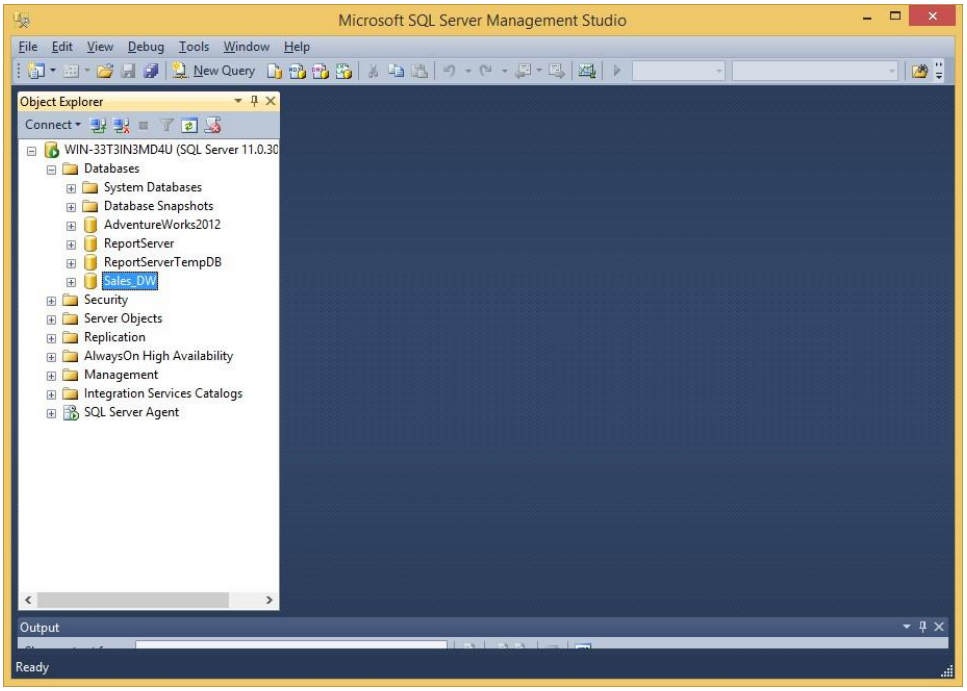
1. **Go to the extracted sql file and double click on it.**

2. **New Sql Query Editor will be opened containing Sales_DW Database**.



3. **Click on execute or press F5 by selecting query one by one or directly click on Execute.**

4. **After completing execution save and close SQL Server Management studio & Reopen to see Sales_DW in Databases Tab.**

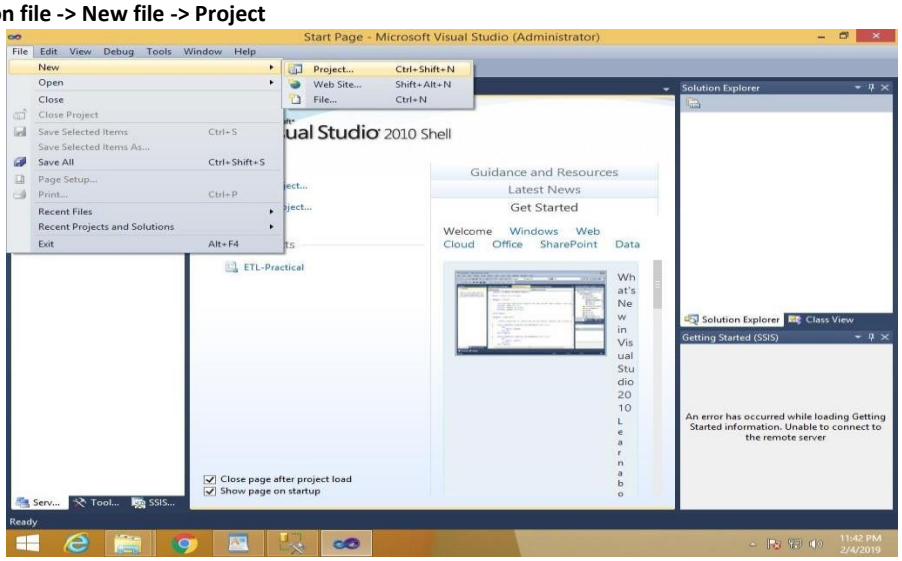


Step 2: Start SSDT environment and create New Data Source

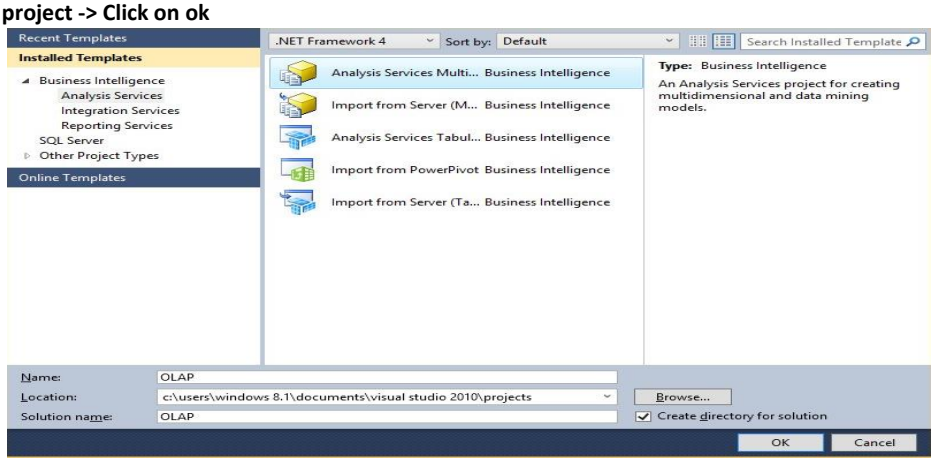
Go to Sql Server Data Tools --> Right click and run as administrator

In Visual studio or Data tools

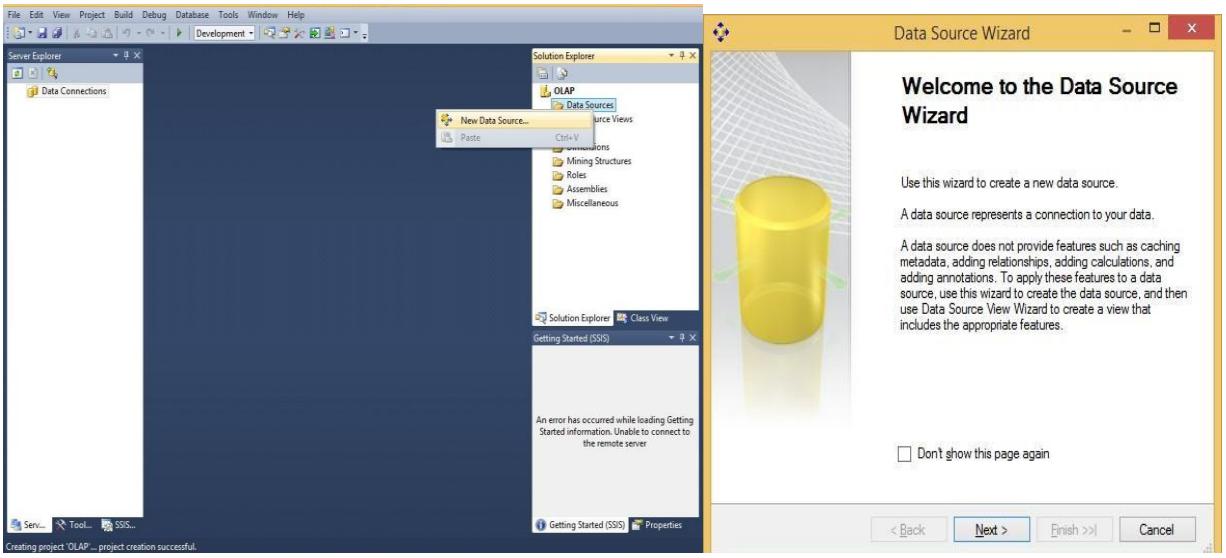
Step 1:- Click on file -> New file -> Project



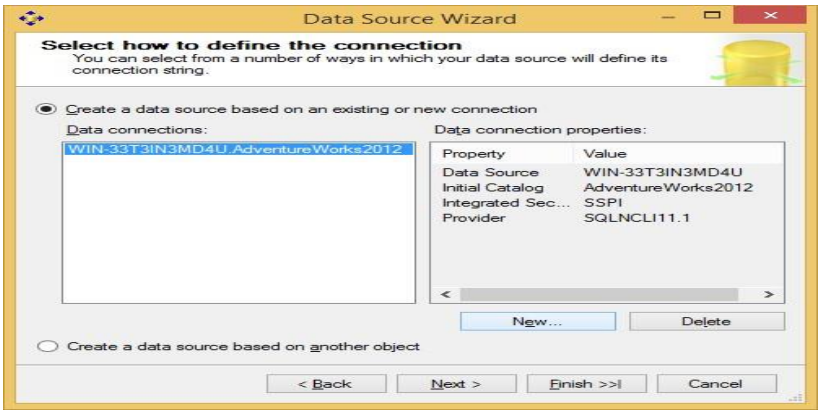
Step 2:- In Business Intelligence -> Select Analysis services Multidimensional and Data Mining models -> Name Your project -> Click on ok



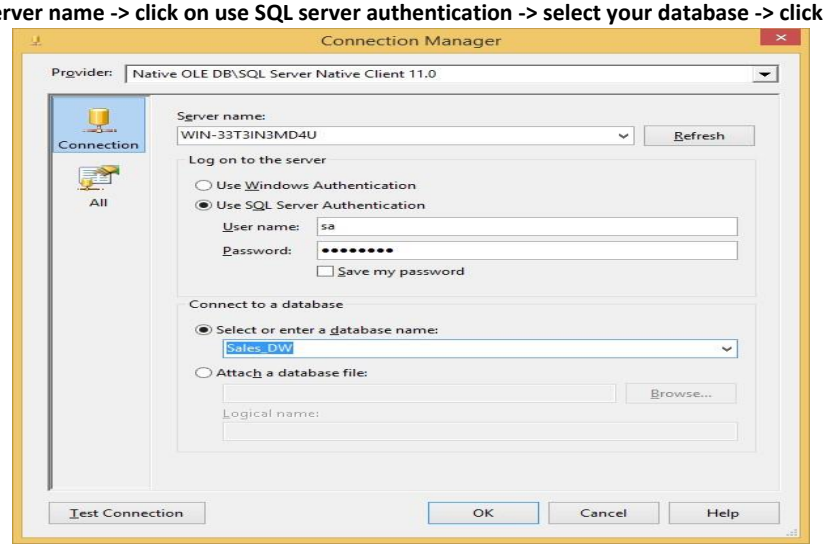
Step 3:- Right click on Data Sources in solution explorer -> New Data Source -> Data Source Wizard appears

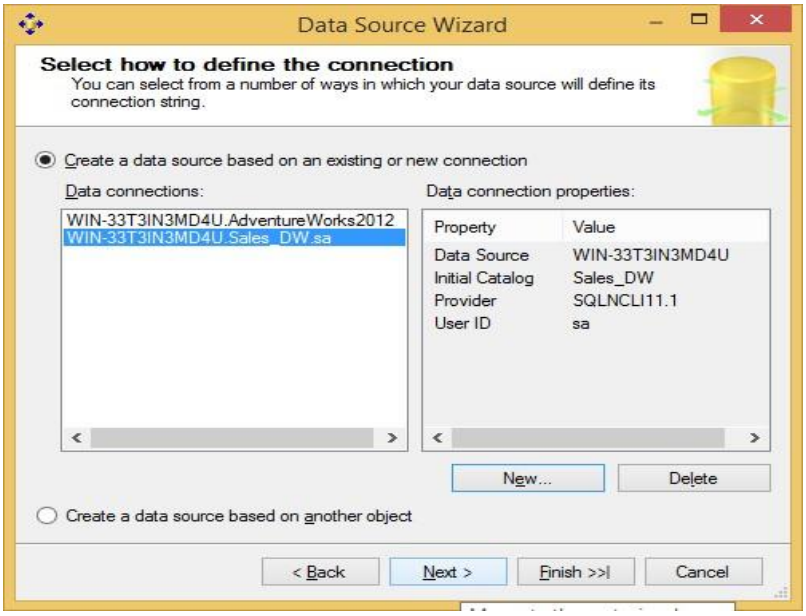


Step 4:- Click on new



Step 5:- Write server name -> click on use SQL server authentication -> select your database -> click on Test connection -> ok

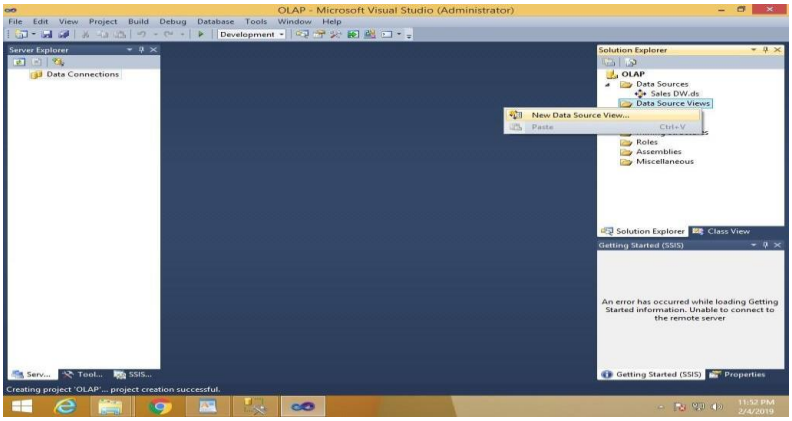




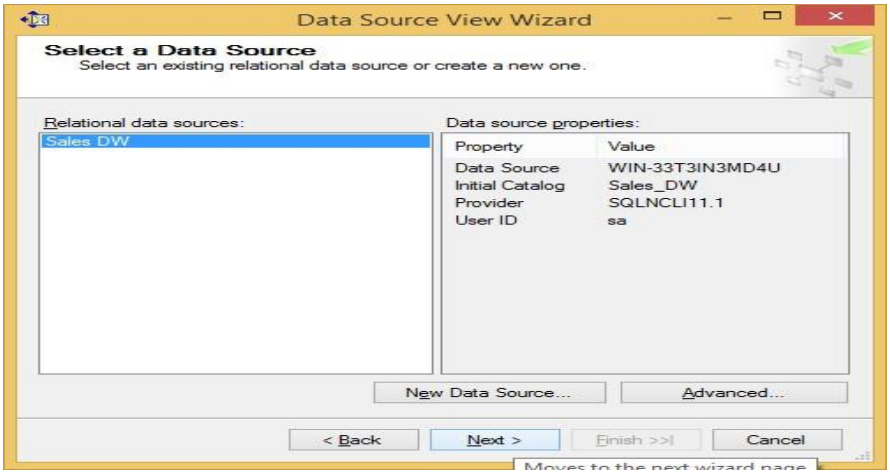
Step 6:- Select Inherit ->Next -> Finish

Step 7:- Sales_DW.ds gets created under Data Sources in Solution Explorer

Step 8:-Creating New Data Source View -> In Solution explorer right click on Data Source View → Select New Data Source View ->Click on next

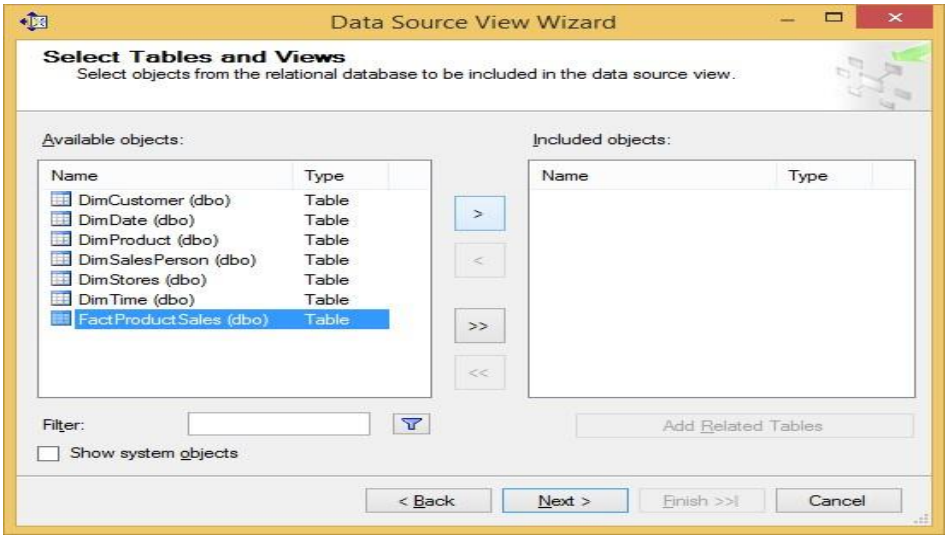


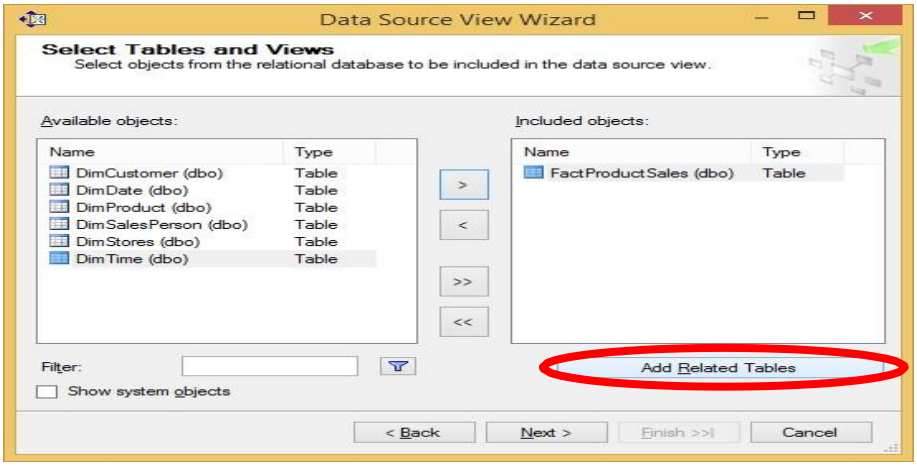
Step 8:- Select Sales_DW -> next

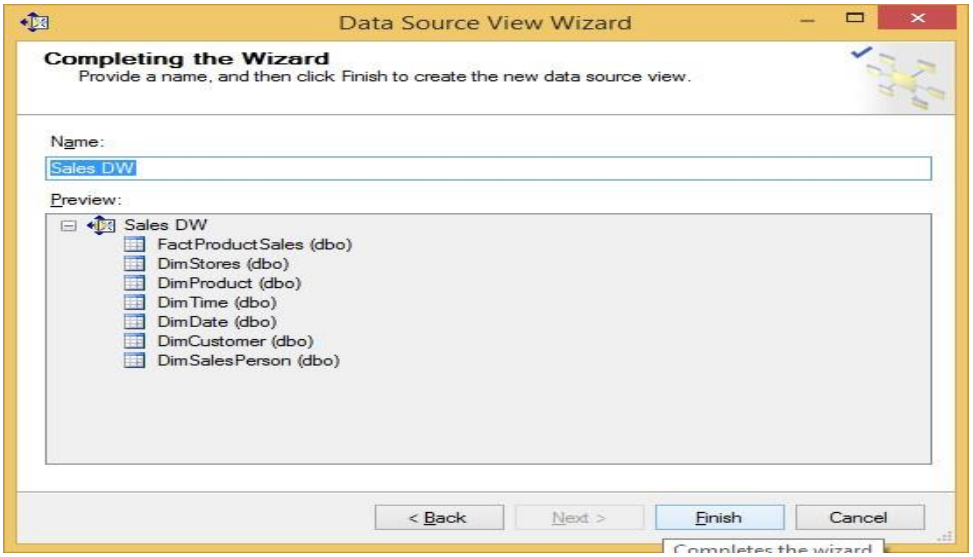


Step 9:- Select FactProductSales(dbo) from Available objects and put in Includes

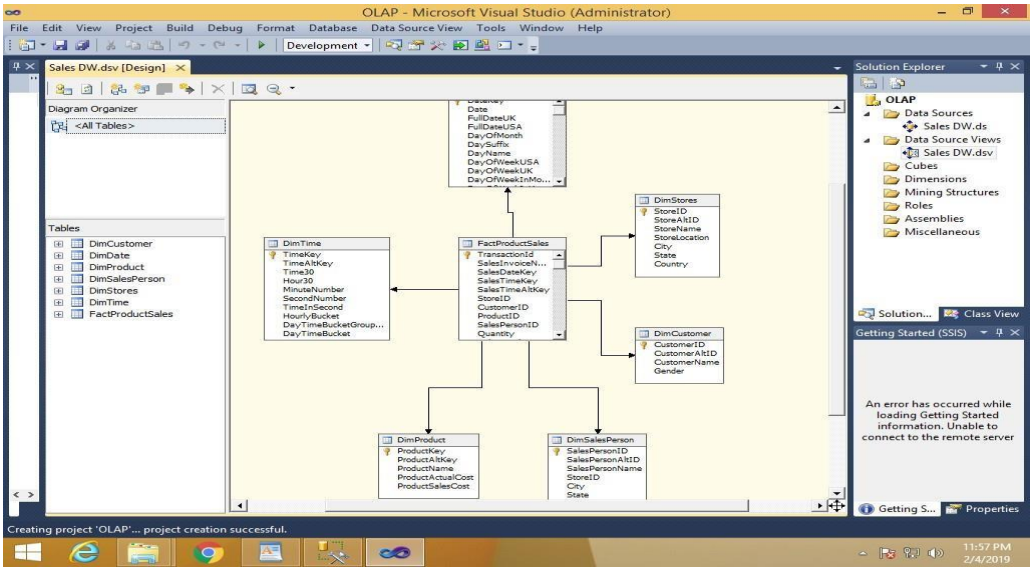
Objects by clicking on  **-> click on add related tables -> Next**


Data Source View Wizard
Select Tables and Views
Select objects from the relational database to be included in the data source view.
Available objects:
Name
Type
DimCustomer (dbo) Table
DimDate (dbo) Table
DimProduct (dbo) Table
DimSalesPerson (dbo) Table
DimStores (dbo) Table
DimTime (dbo) Table
FactProductSales (dbo) Table
Included objects:
Name
Type
>
<
>>
<<
Filter:
Add Related Tables
Show system objects
< Back
Next >
Finish >>|
Cancel

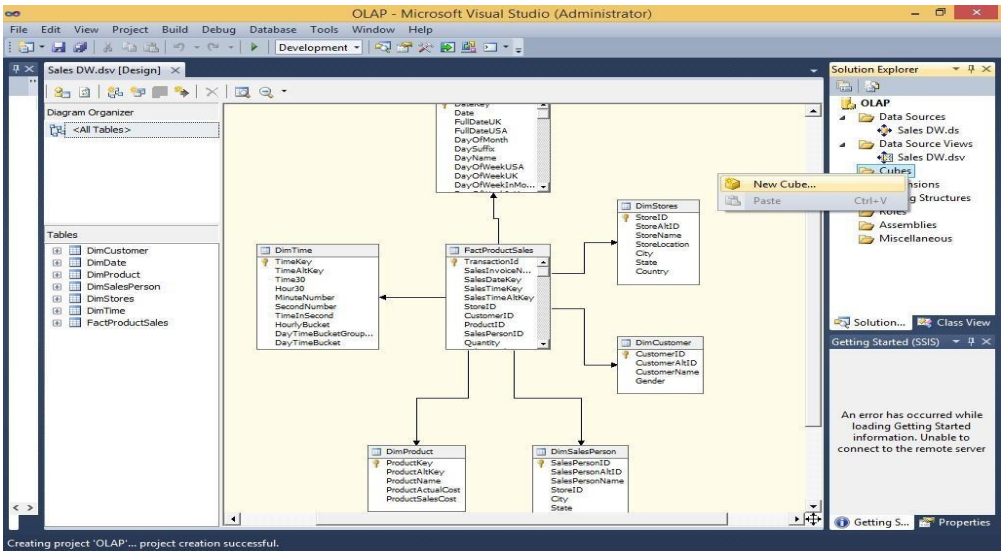

Data Source View Wizard
Select Tables and Views
Select objects from the relational database to be included in the data source view.
Available objects:
Name
Type
DimCustomer (dbo) Table
DimDate (dbo) Table
DimProduct (dbo) Table
DimSalesPerson (dbo) Table
DimStores (dbo) Table
DimTime (dbo) Table
Included objects:
Name
Type
FactProductSales (dbo) Table
>
<
>>
<<
Filter:
Add Related Tables
Show system objects
< Back
Next >
Finish >>|
Cancel


Data Source View Wizard
Completing the Wizard
Provide a name, and then click Finish to create the new data source view.
Name:
Sales DW
Preview:
Sales DW
FactProductSales (dbo)
DimStores (dbo)
DimProduct (dbo)
DimTime (dbo)
DimDate (dbo)
DimCustomer (dbo)
DimSalesPerson (dbo)
< Back
Next >
Finish
Cancel

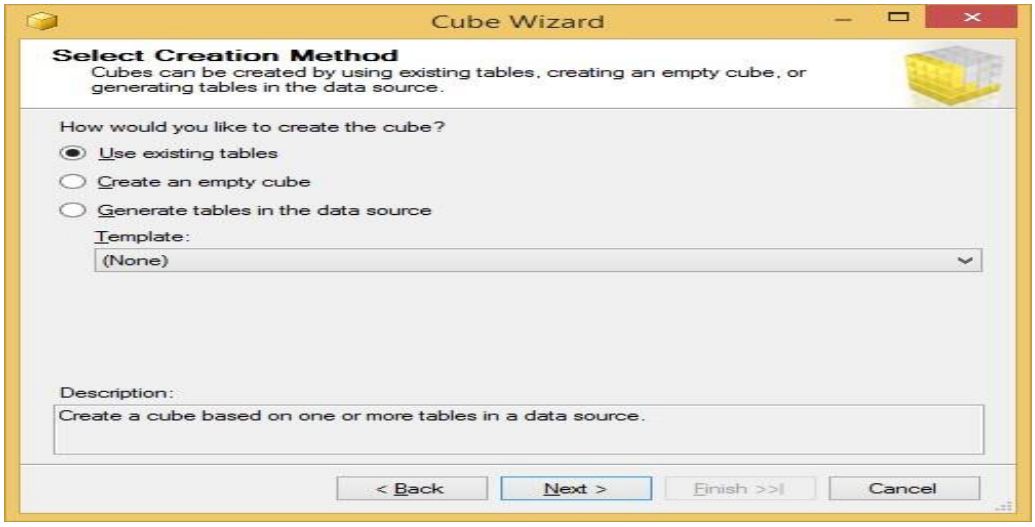
Step 10:- Sales DW.dsv appears in Data Source Views in Solution Explorer.



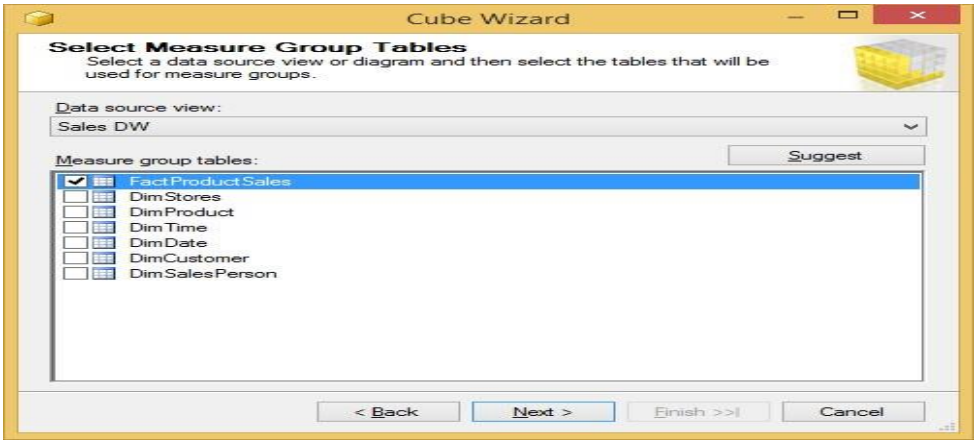
Step 11 : Creating new cube -> Right click on Cubes New Cube

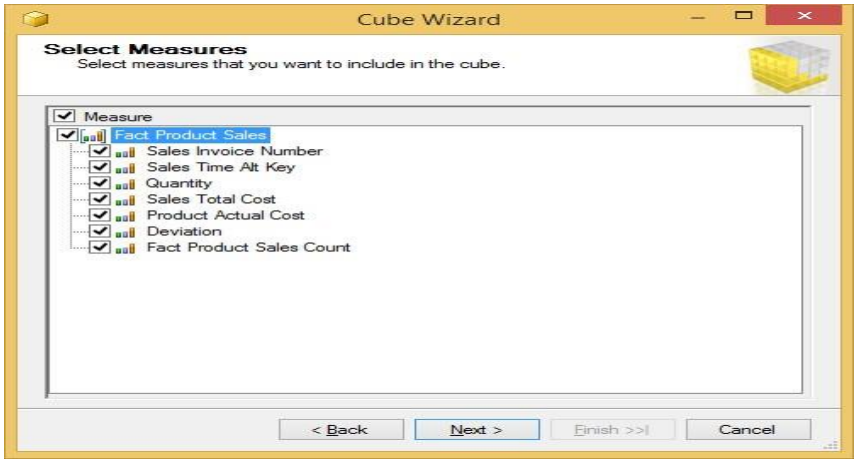


Step 12:-Select Use existing tables in Select Creation Method → Next

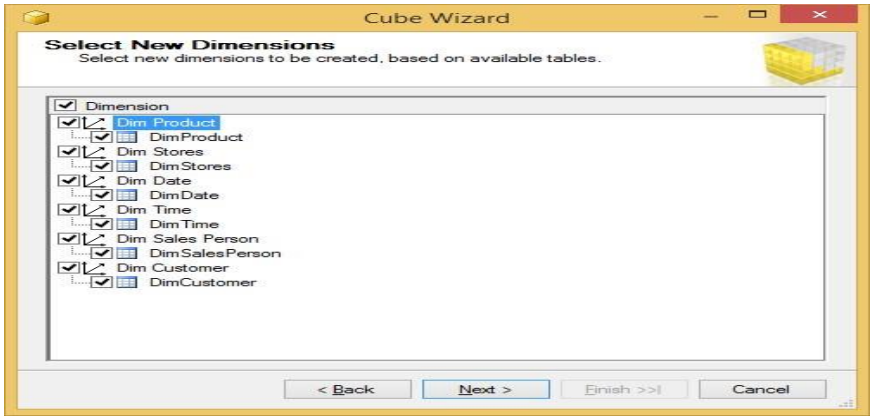


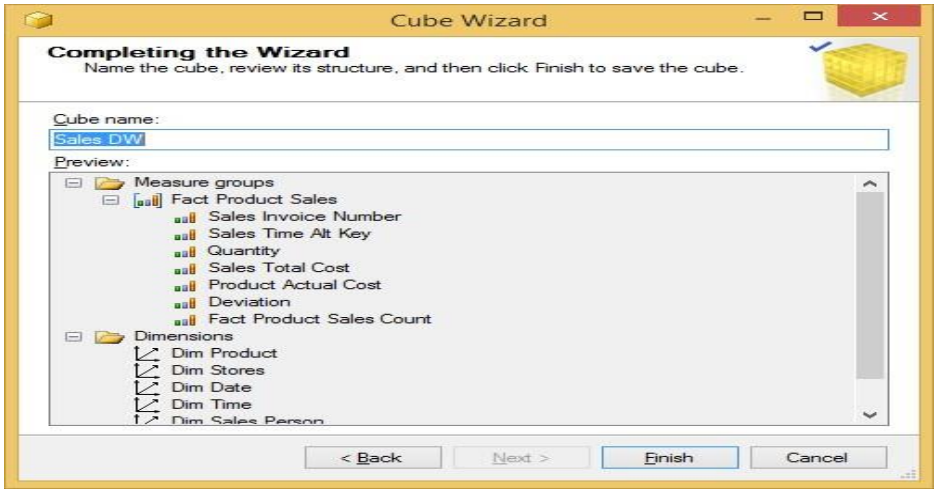
Step 13:-In Select Measure Group Tables Select FactProductSales Click Next -> In Select Measures check all measures Next

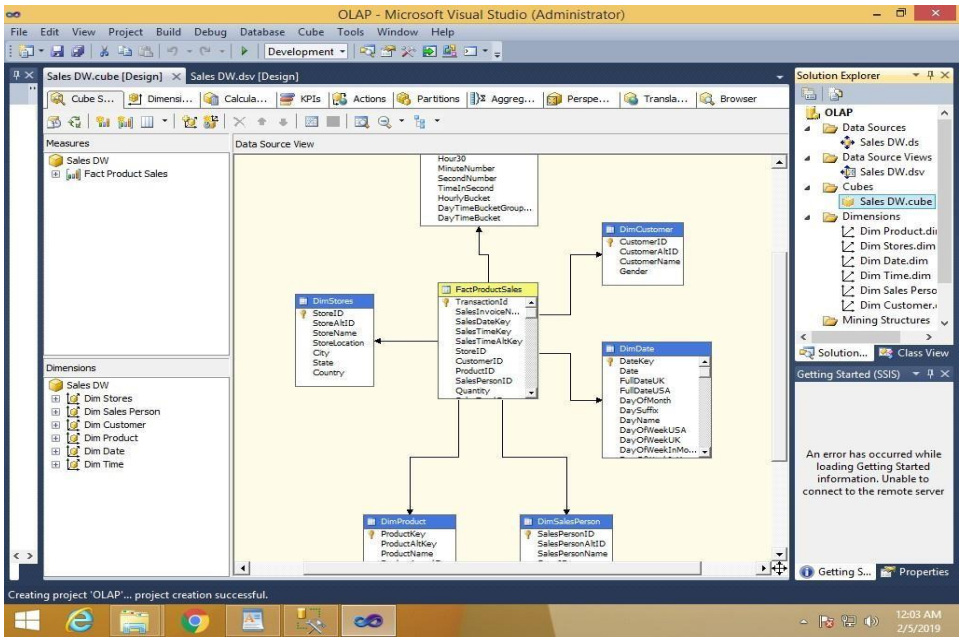




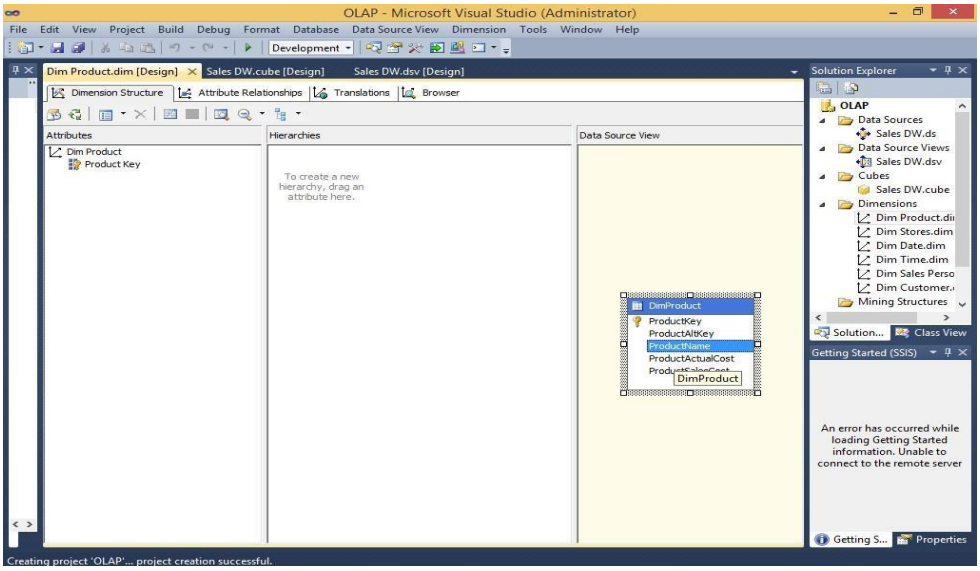
Step 14:- In Select New Dimensions Check all Dimensions->Next -> Click on Finish -> Sales_DW.cube is created

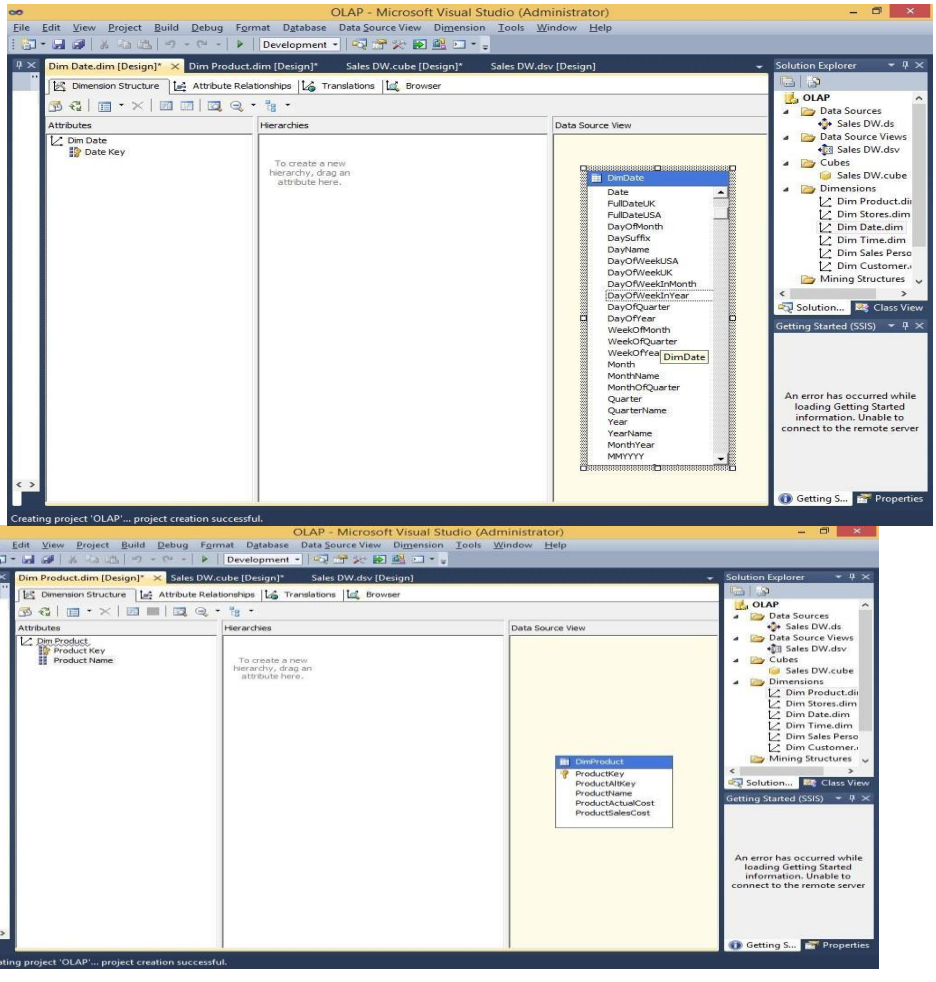




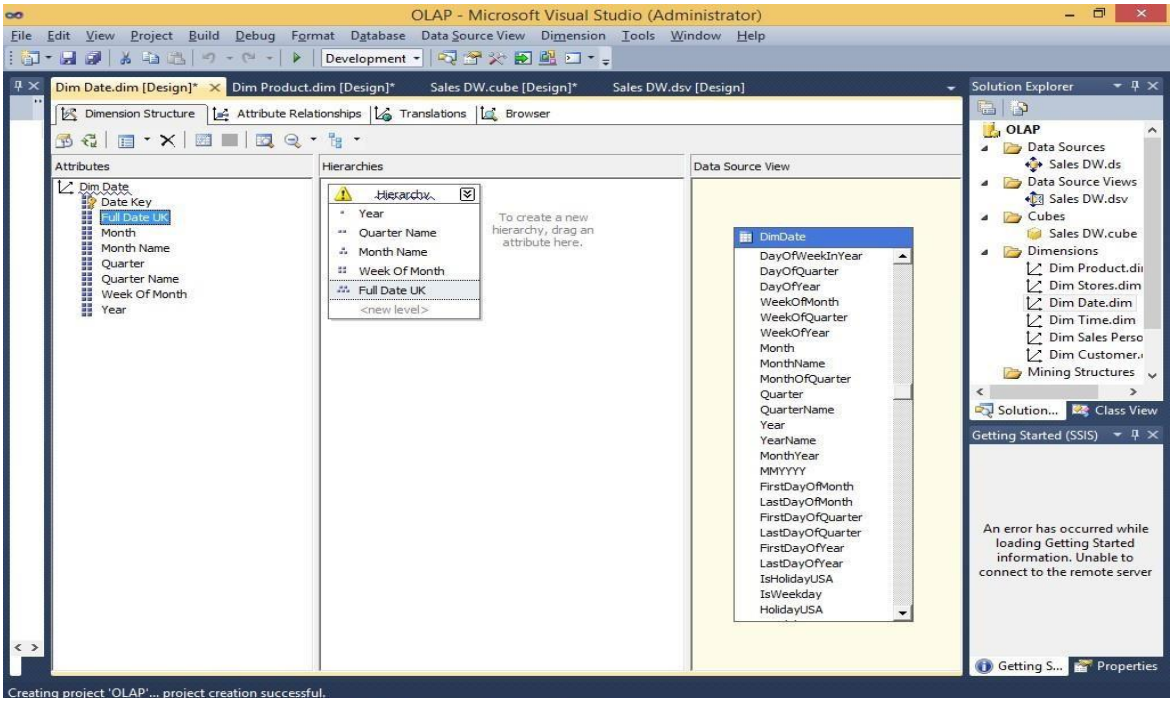


Step 15 : Dimension Modification -> In dimension tab Double Click Dim Product.dim and Drop Product Name from Table in Data Source View and Add in Attribute Pane at left side

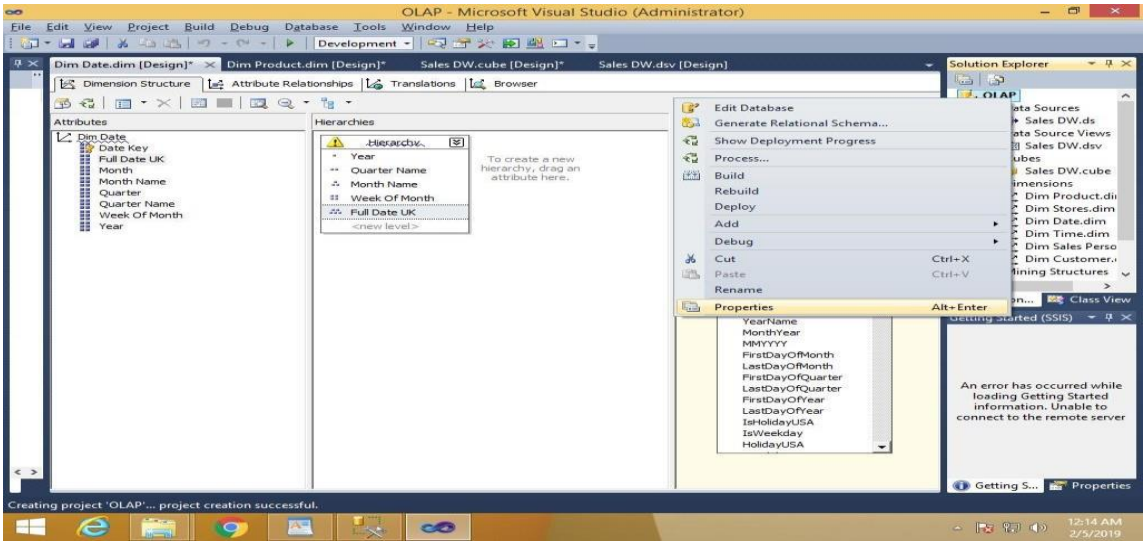


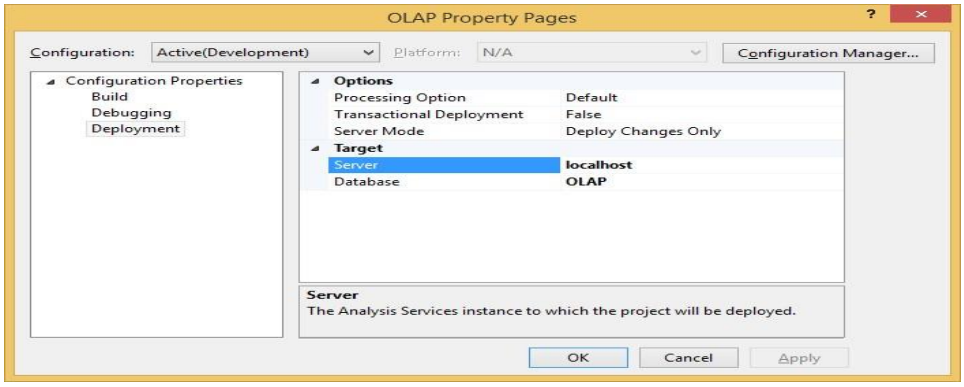


Step 16: Creating Attribute Hierarchy in Date Dimension Double click On Dim Date dimension -> Drag and Drop Fields from Table shown in Data Source View to Attributes-> Drag and Drop attributes from leftmost pane of attributes to middle pane of Hierarchy. Drag fields in sequence from Attributes to Hierarchy window (Year, Quarter Name, Month Name, Week of the Month, Full Date UK)

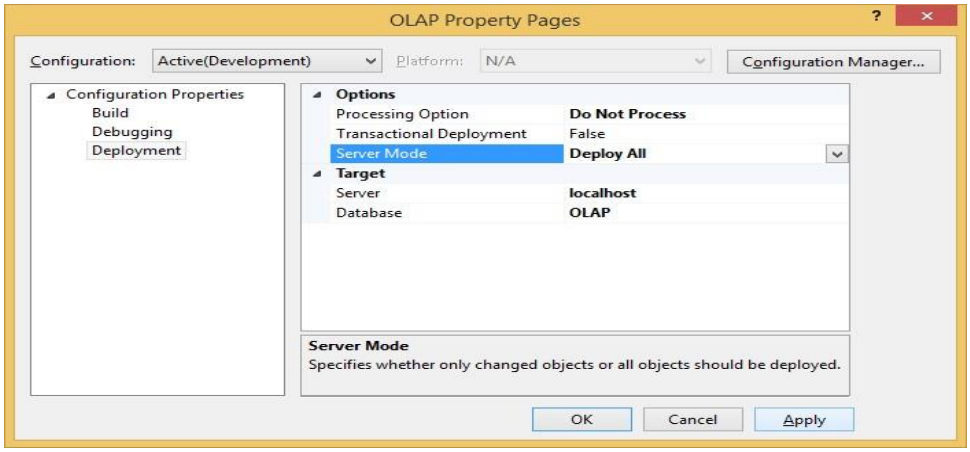


Step 17: Deploy Cube Right click on Project name Properties

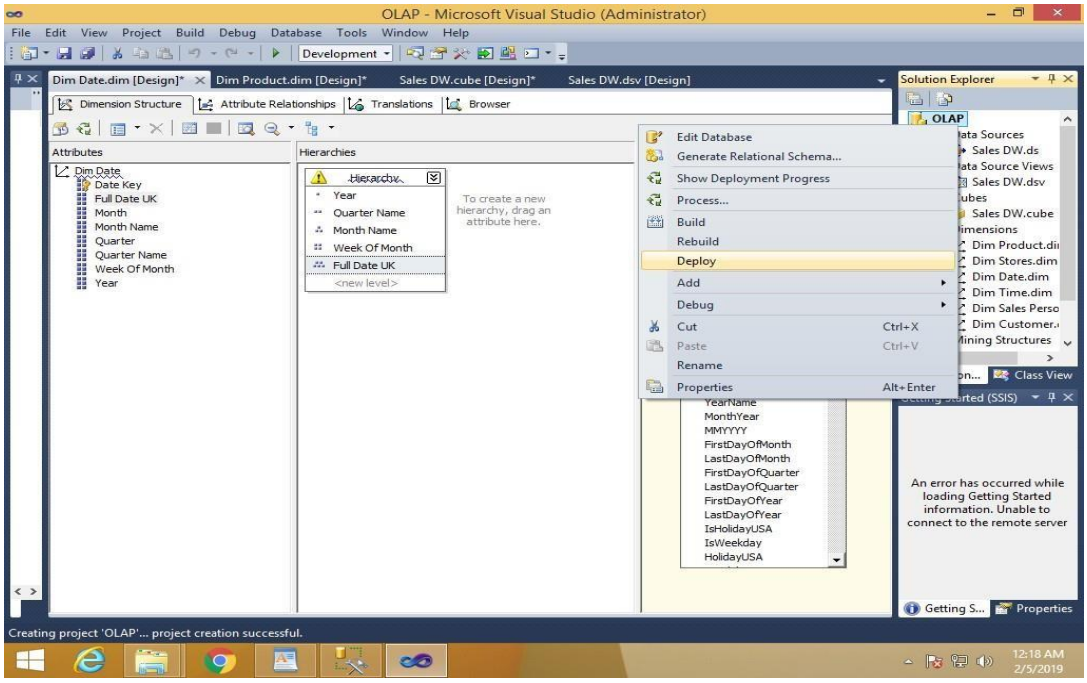


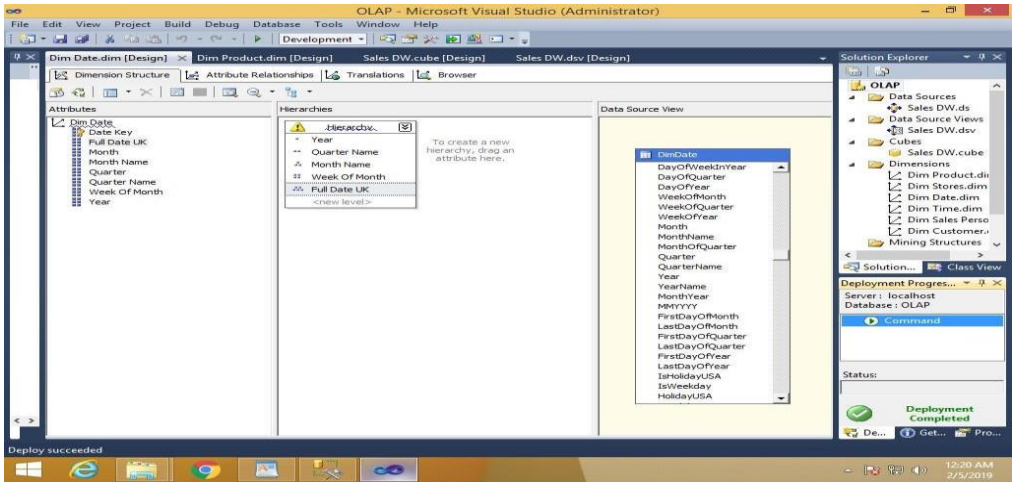


Do following changes and click on Apply & ok

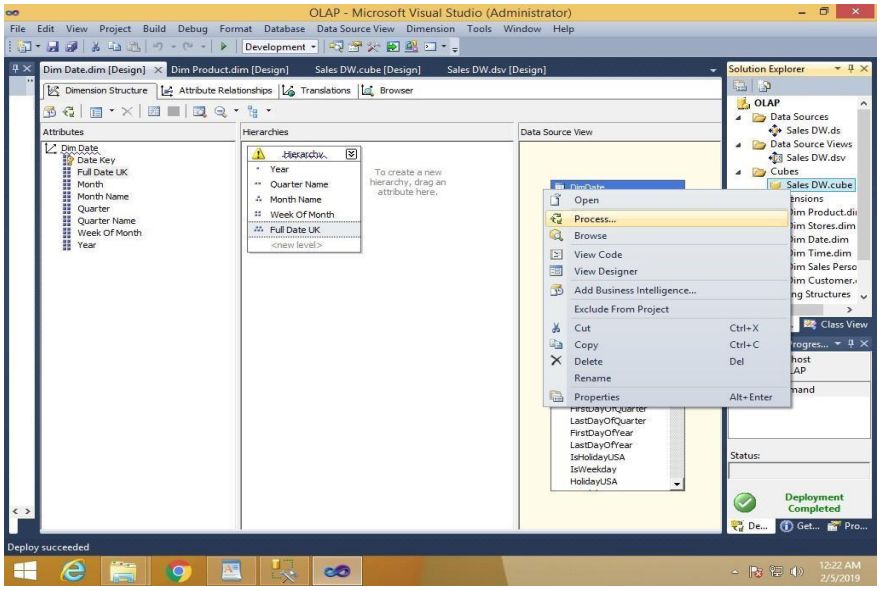


Step 18:-Right click on project name Deploy -> Deployment successful

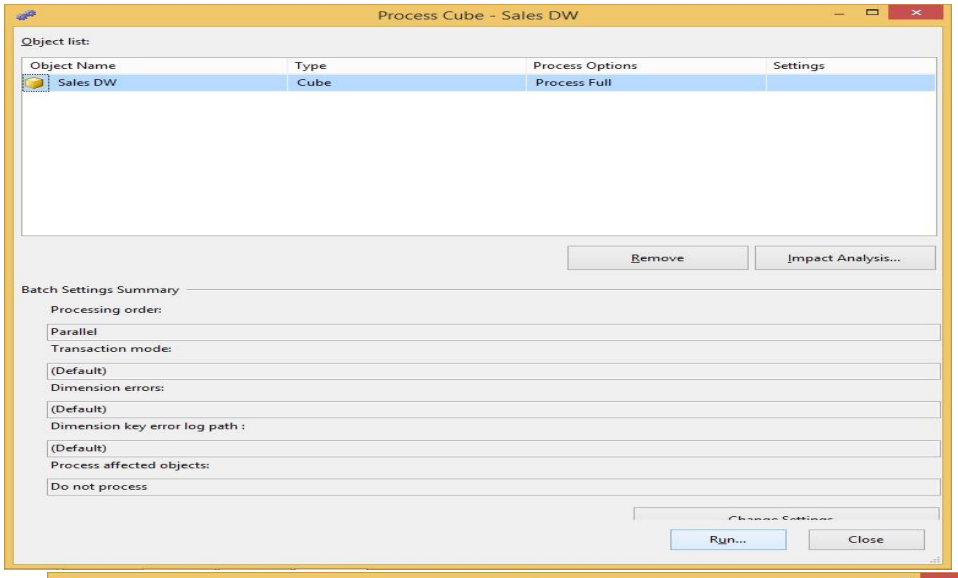


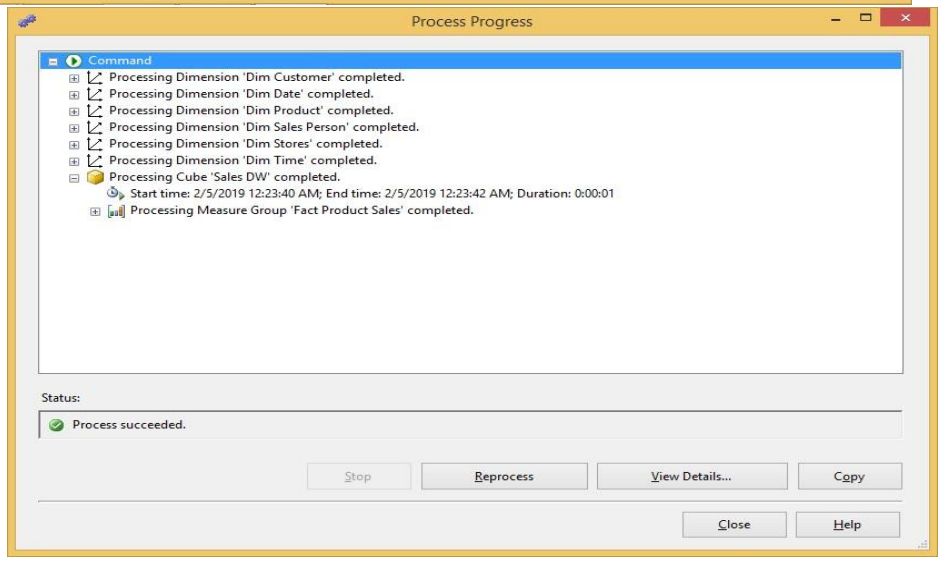


Step 19:- To process cube right click on Sales_DW.cube Process

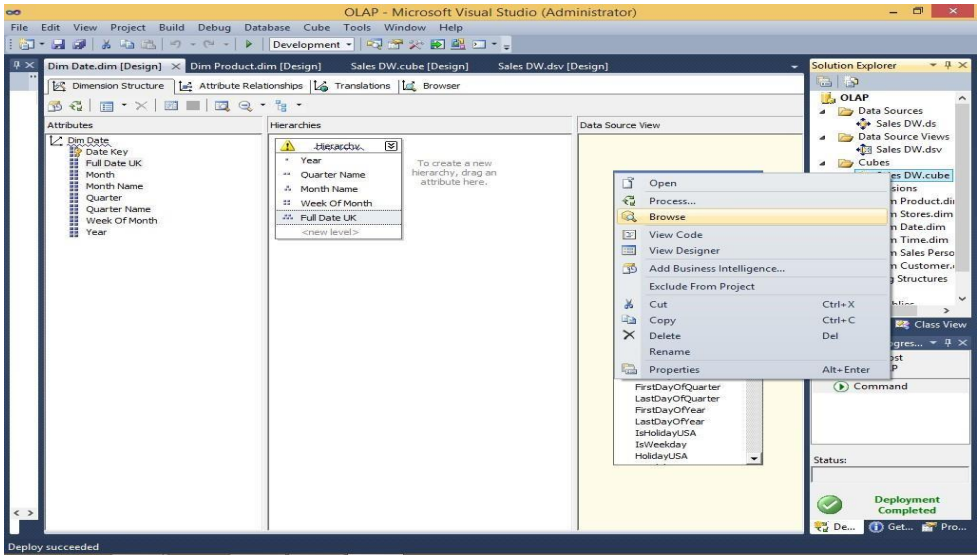

OLAP - Microsoft Visual Studio (Administrator)
File Edit View Project Build Debug Format Database Data Source View Dimension Tools Window Help
Development
Dim Date.dim [Design]
Dim Product.dim [Design]
Sales DW.cube [Design]
Sales DW.dsv [Design]
Dimension Structure
Attribute Relationships
Translations
Browser
Attributes
Dim Date
Date Key
Full Date UK
Month
Month Name
Quarter
Quarter Name
Week Of Month
Year
Hierarchies
Hierarchy
Year
Quarter Name
Month Name
Week Of Month
Full Date UK
<new level>
To create a new hierarchy, drag an attribute here.
Data Source View
Open
Process...
Browse
View Code
View Designer
Add Business Intelligence...
Exclude From Project
Cut Ctrl+X
Copy Ctrl+C
Delete Del
Rename
Properties Alt+Enter
Solution Explorer
OLAP
Data Sources
Sales DW.ds
Data Source Views
Sales DW.dsv
Cubes
Sales DW.cube
Class View
Status:
Deployment Completed
Deploy succeeded

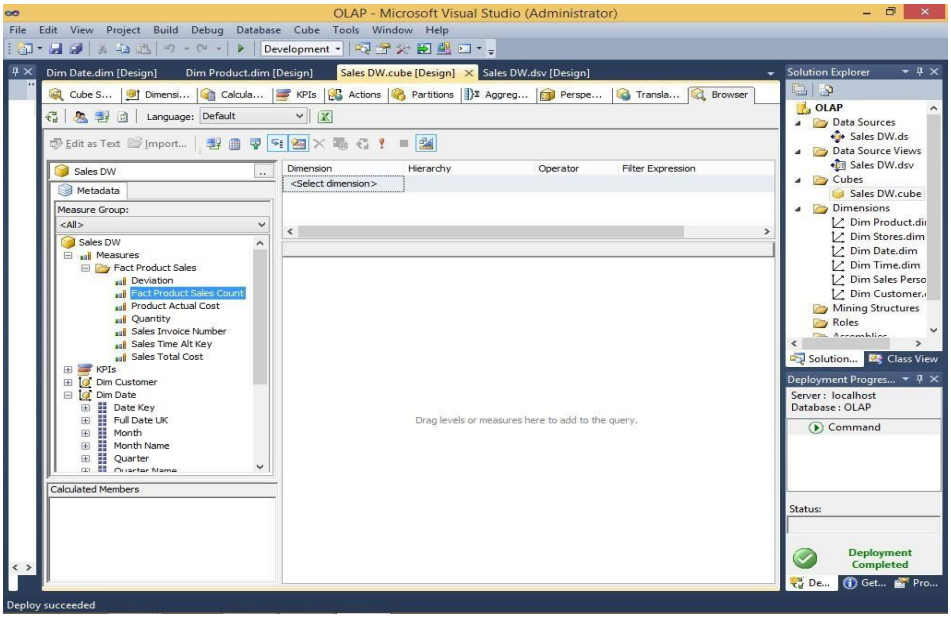
Click on run


Process Cube - Sales DW
Object list:
Object Name
Type
Process Options
Settings
Sales DW
Cube
Process Full
Remove
Impact Analysis...
Batch Settings Summary
Processing order:
Parallel
Transaction mode:
(Default)
Dimension errors:
(Default)
Dimension key error log path :
(Default)
Process affected objects:
Do not process
Run...
Close


Process Progress
Command
Processing Dimension 'Dim Customer' completed.
Processing Dimension 'Dim Date' completed.
Processing Dimension 'Dim Product' completed.
Processing Dimension 'Dim Sales Person' completed.
Processing Dimension 'Dim Stores' completed.
Processing Dimension 'Dim Time' completed.
Processing Cube 'Sales DW' completed.
Start time: 2/5/2019 12:23:40 AM; End time: 2/5/2019 12:23:42 AM; Duration: 0:00:01
Processing Measure Group 'Fact Product Sales' completed.
Status:
Process succeeded.
Stop
Reprocess
View Details...
Copy
Close
Help

Step 20->Browse the cube for analysis in solution explorer





PRACTICAL 8

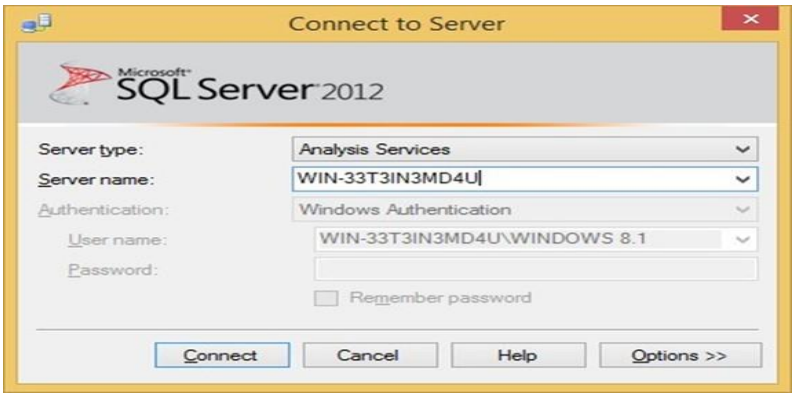
AIM:Execute the MDX queries to extract the data from the data warehouse.

Step 1: Open SQL Server Management Studio and connect to Analysis Services.

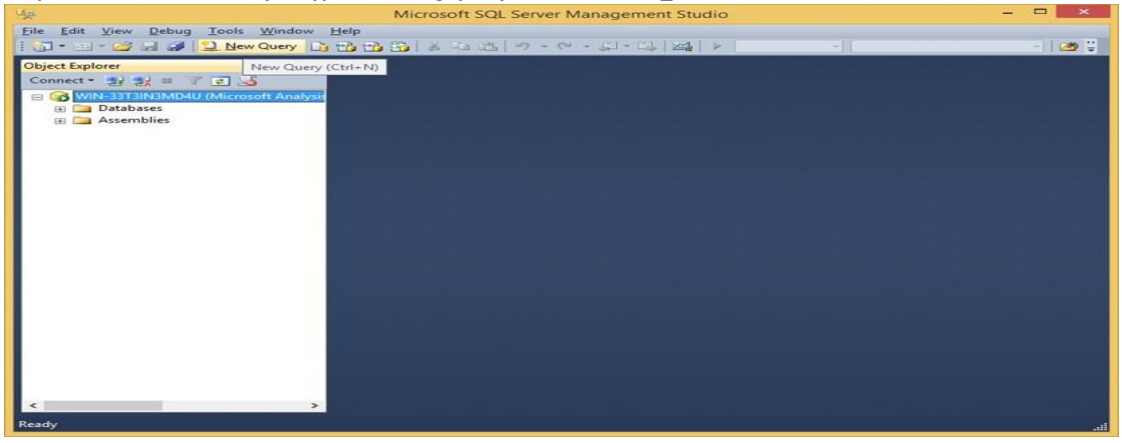
Server type: Analysis Services

Server Name: (according to base machine)

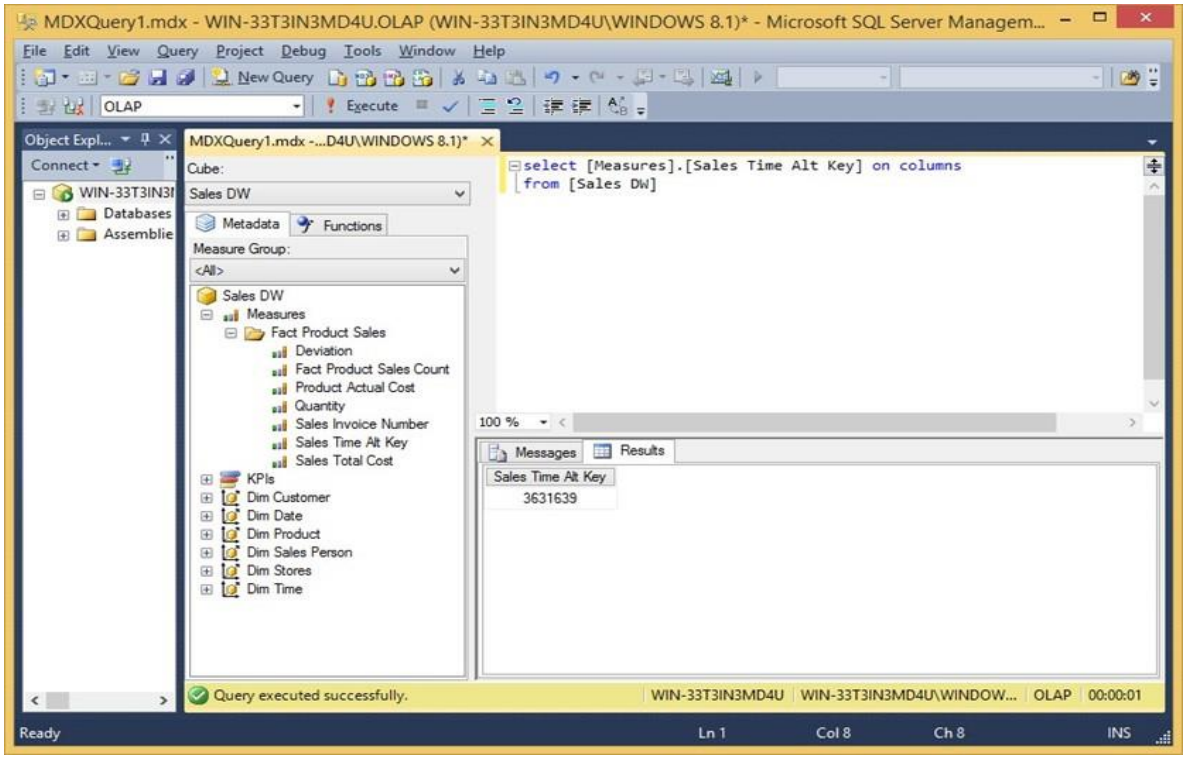
Click on connect



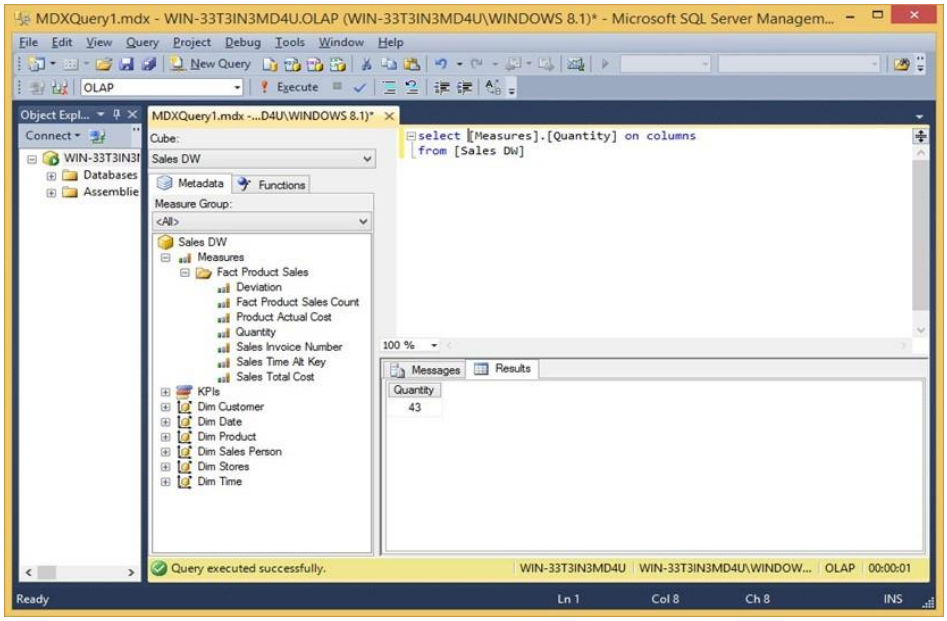
Step 2: Click on New Query & type following query based on Sales_DW



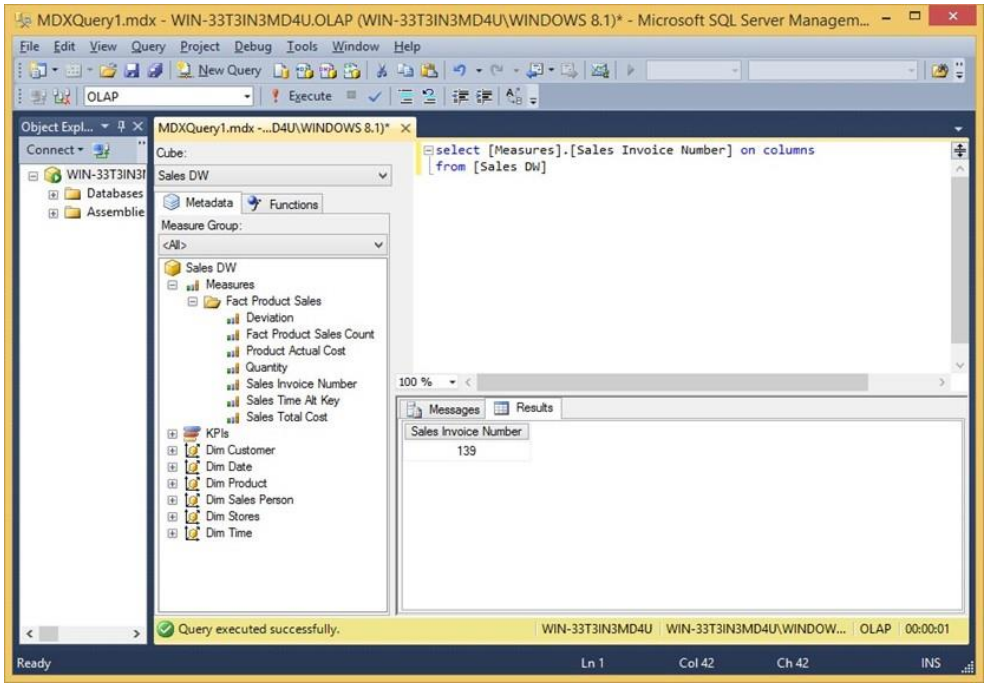
select [Measures].[Sales Time Alt Key] on columns from [Sales DW] Click on execute



select [Measures].[Quantity] on columns from [Sales DW]

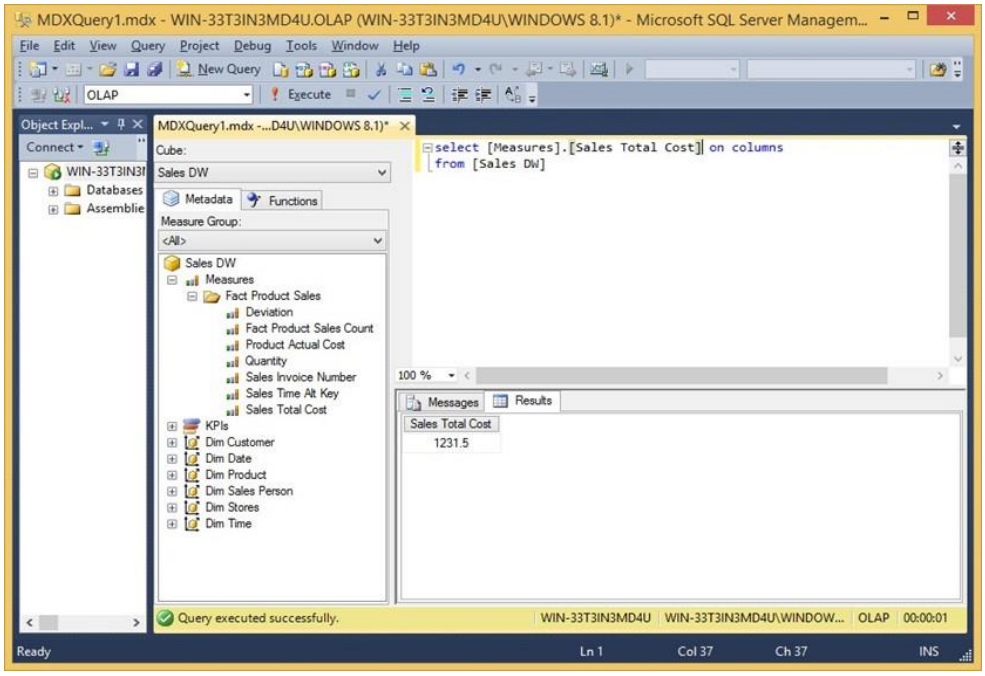


select [Measures].[Sales Invoice Number] on columns from [Sales DW]



select [Measures].[Sales Total Cost] on columns from

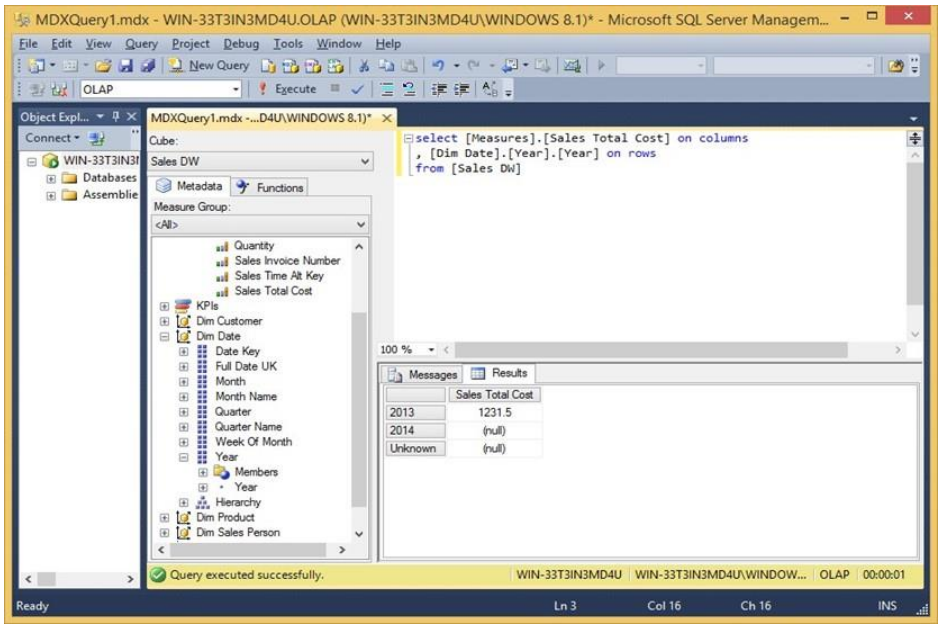
[Sales DW]



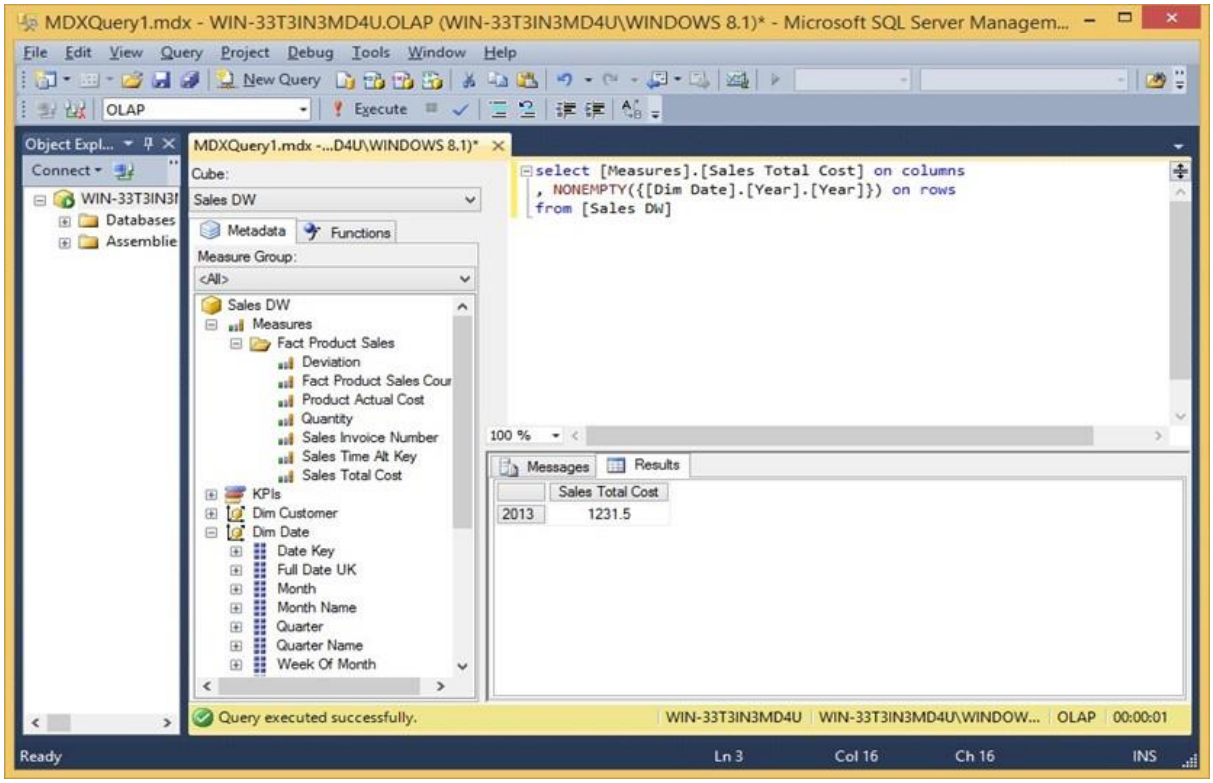
select [Measures].[Sales Total Cost] on columns

, [Dim Date].[Year].[Year] on rows from

[Sales DW]



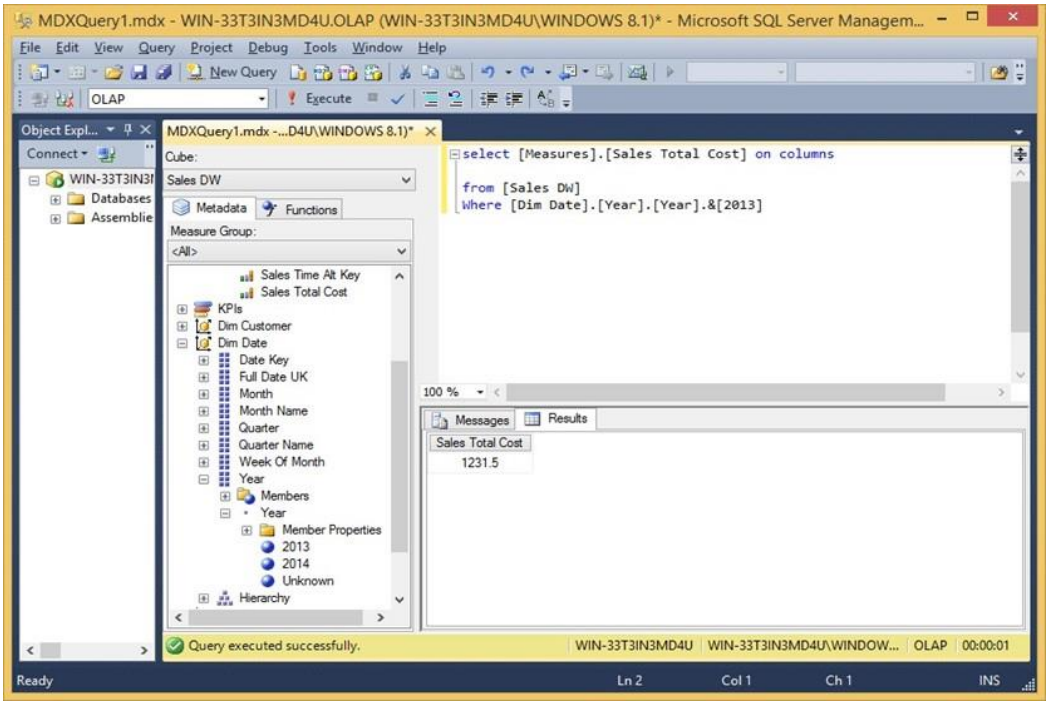
select [Measures].[Sales Total Cost] on columns , NONEMPTY({[Dim Date].[Year].[Year]}) on rows from [Sales DW]



select [Measures].[Sales Total Cost] on columns from

[Sales DW]

Where [Dim Date].[Year].[Year].&[2013]



MDX query

IF error occur

IN Data tools

Step 1 :- In cubes , right click on Sales_DW.cubes -> select process ->in object list copy error

Step 2 :- In MSSMS go to Data engine -> click on security -> right click on login in -> new login -> go to general in login name go to search -> paste the error -> click on ok

Step 3 :- go to server Roles -> select sysadmin -> click on ok

PRACTICAL 9

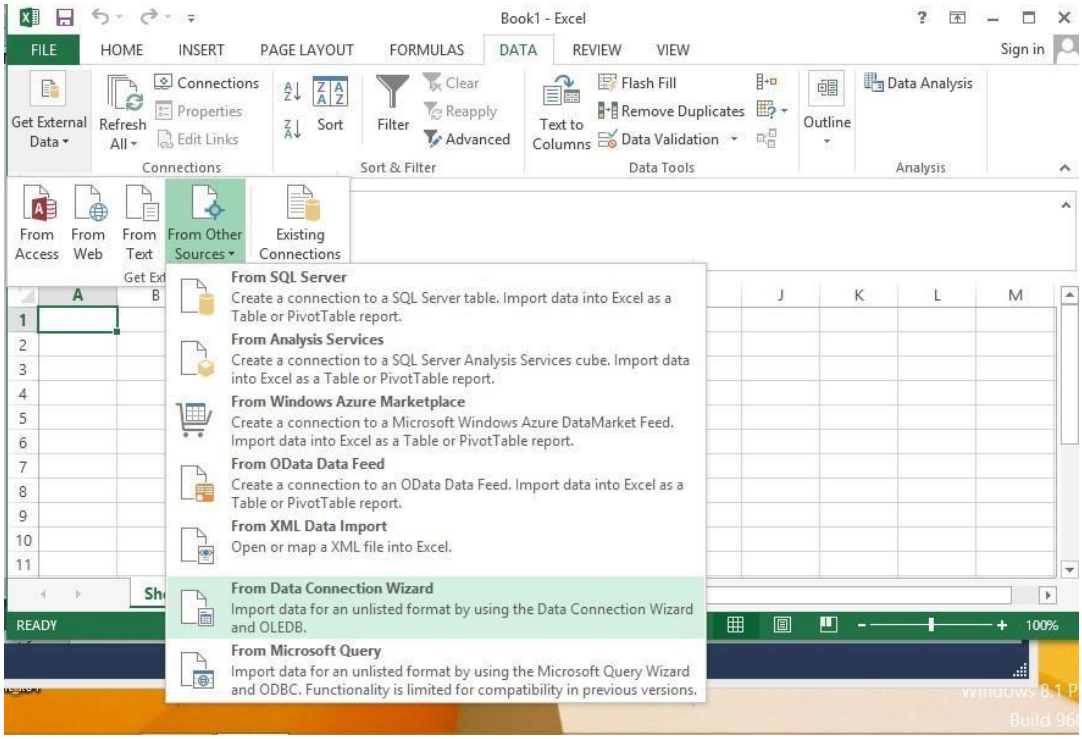
AIM: Import the data warehouse data in Microsoft Excel and create the Pivot table and Pivot Chart

(Ms Office Professional is used to make sure Power View is enabled for visualisation.)

Step 1: Open Excel 2013 (Professional)

Go to Data tab → Get External Data → From Other Sources → From

Data Connection Wizard



Step 2: In Data Connection Wizard → Select Microsoft SQL Server →

Click on Next

Step 3: In connect to Database Server provide Server name(Microsoft

SQL Server Name)

Provide password for sa account as given during installation of SQL

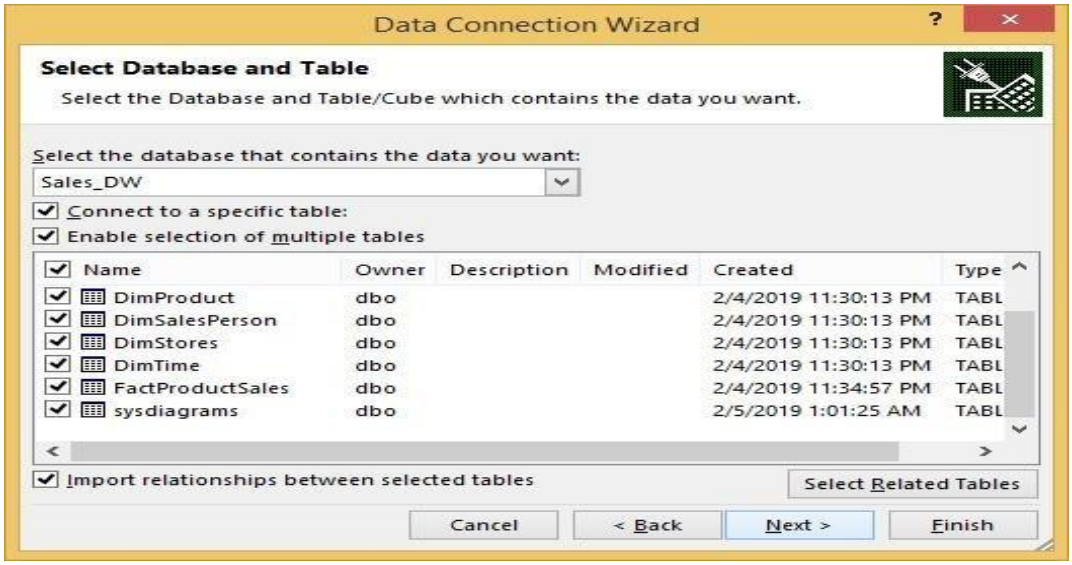
Server 2012 full version)

Password: admin123

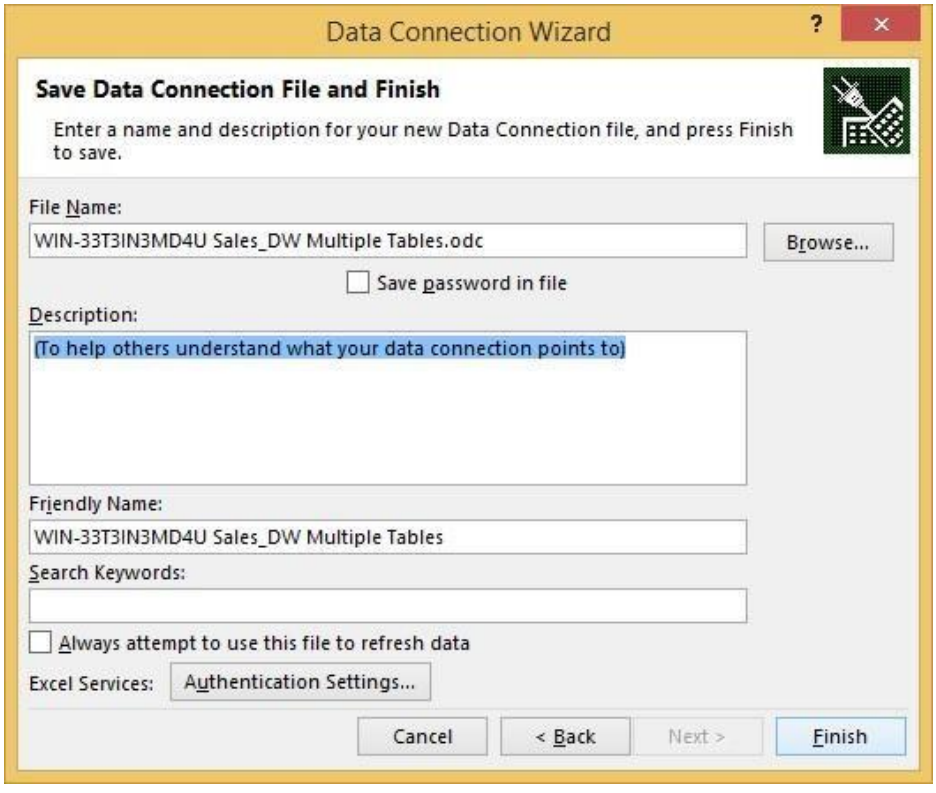
Click on Next

Step 4: In Select Database and Table → Select Sales_DW (already created in SQL) → check all dimensions and import

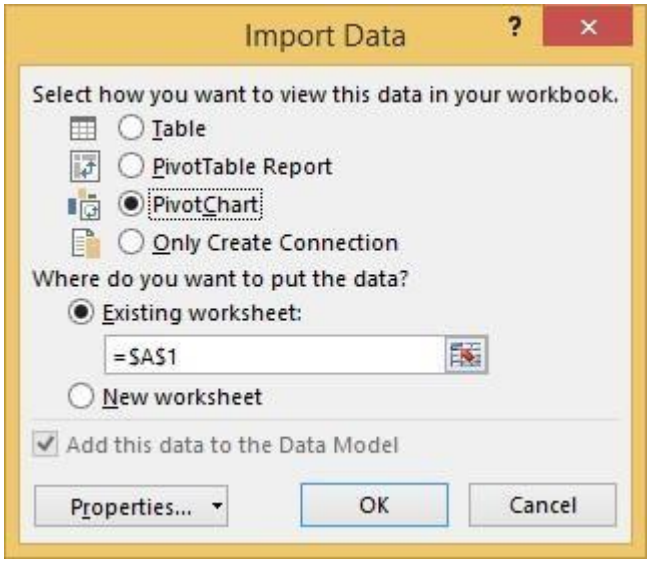
relationships between selected tables



Step 5: In save data connection files browse path and click on Finish

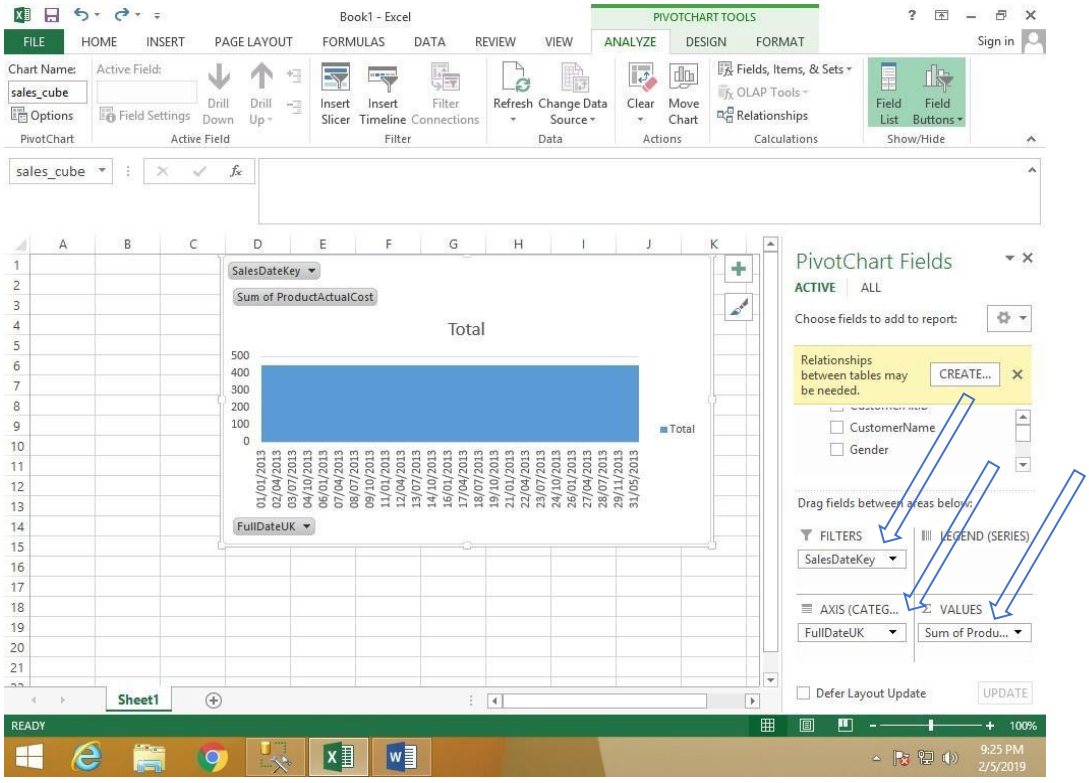


Step 6: In import data select Pivot Chart and click on OK

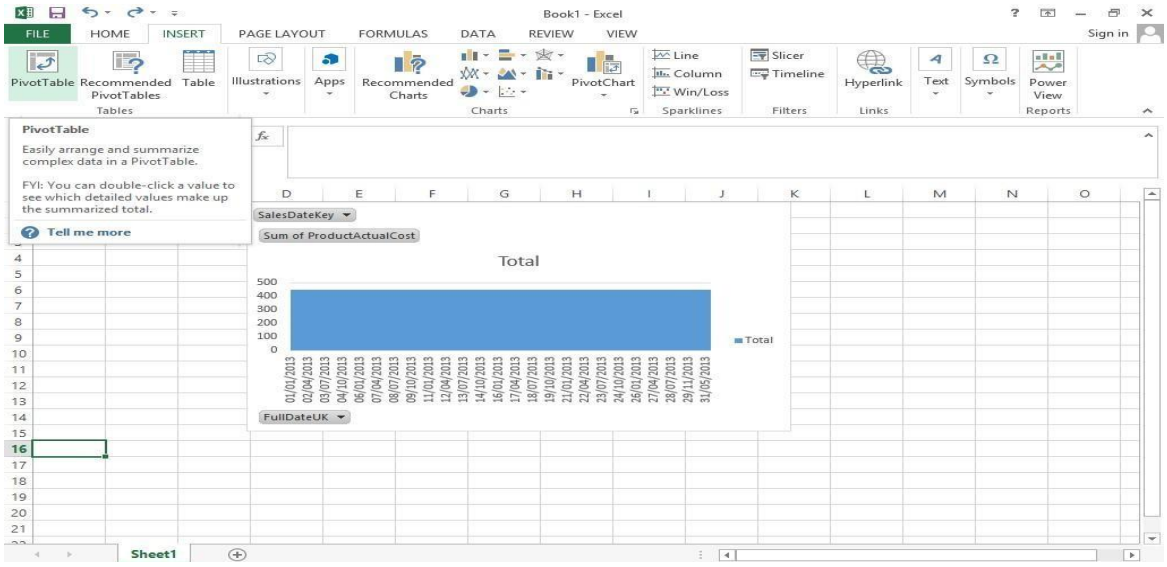


Step 7: In fields put SalesDateKey in filters, FullDateUK in axis and Sum of

ProductActualCost in values

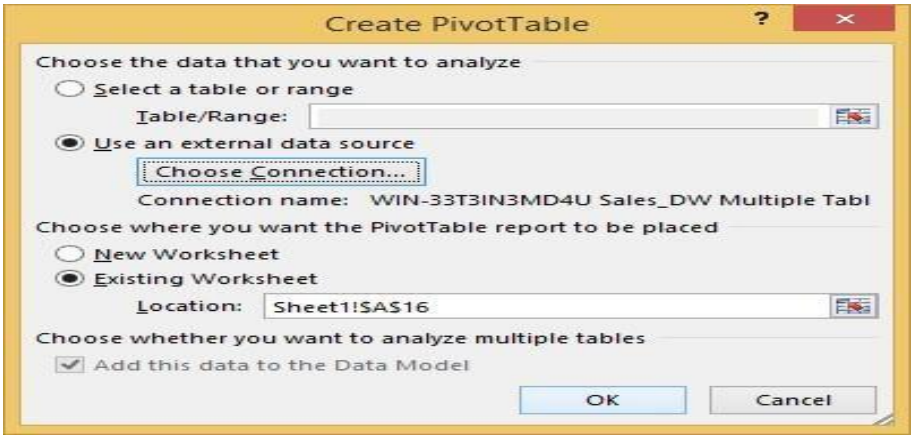


Step 8: In Insert Tab → go to Pivot Table

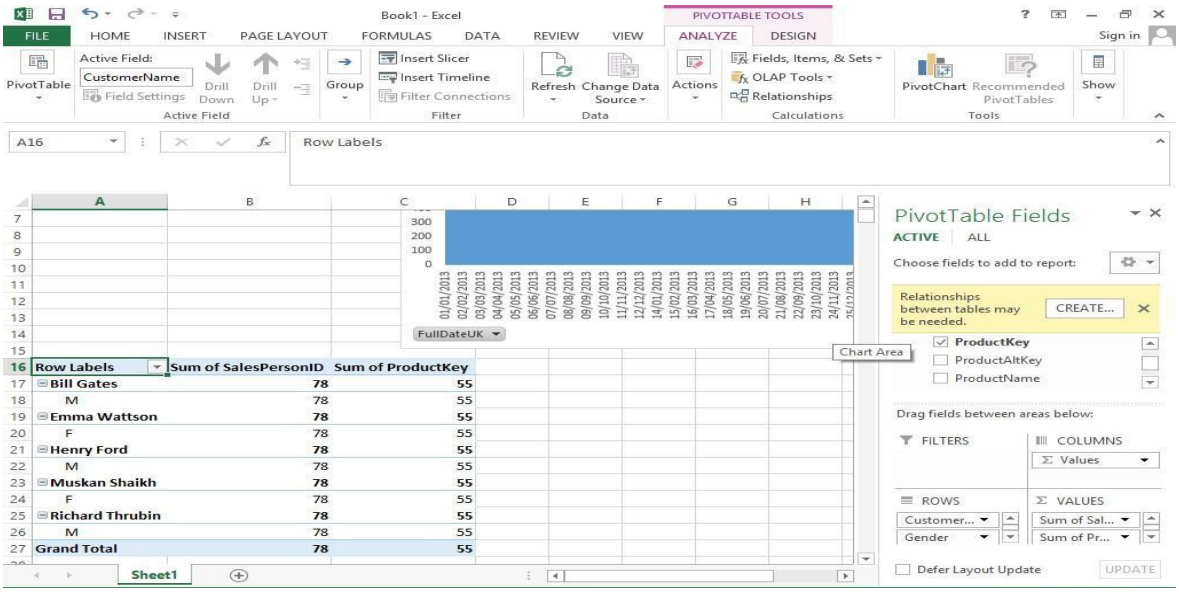


Step 9: Click on Choose Connection to select existing connection with

Sales_DW and click on open



Pivot table and Pivot chart is created



Row Labels	Sum of SalesPersonID	Sum of ProductKey
Bill Gates	78	55
M	78	55
Emma Wattson	78	55
F	78	55
Henry Ford	78	55
M	78	55
Muskan Shaikh	78	55
F	78	55
Richard Thrubin	78	55
M	78	55
Grand Total	78	55

Practical 10

AIM:-Perform data analysis using what if analysis.

Steps1:-Go to excel→Create a table→add items and their costs.

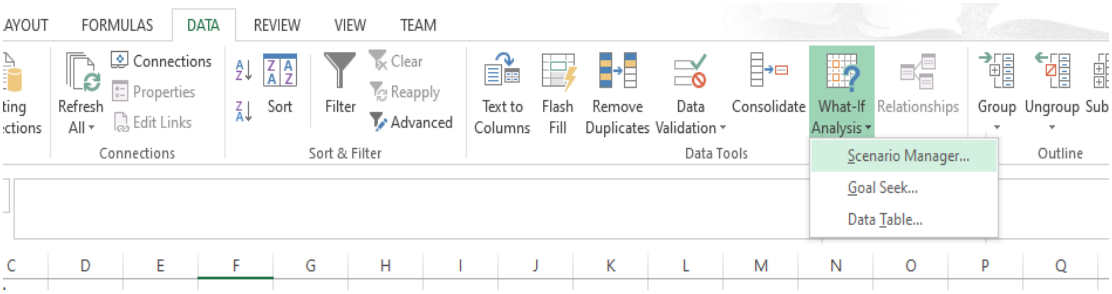
	A	B	C
1	SR.NO	Items	Costs
2	1	Machinery	60000
3	2	Carriage	8000
4	3	Transport	30000
5	4	Office Equipments	6000
6	5	Postage	7000
7	6	Miscalleneous	3000
8	7	Generator	5000
9		Total Cost	119000
10			

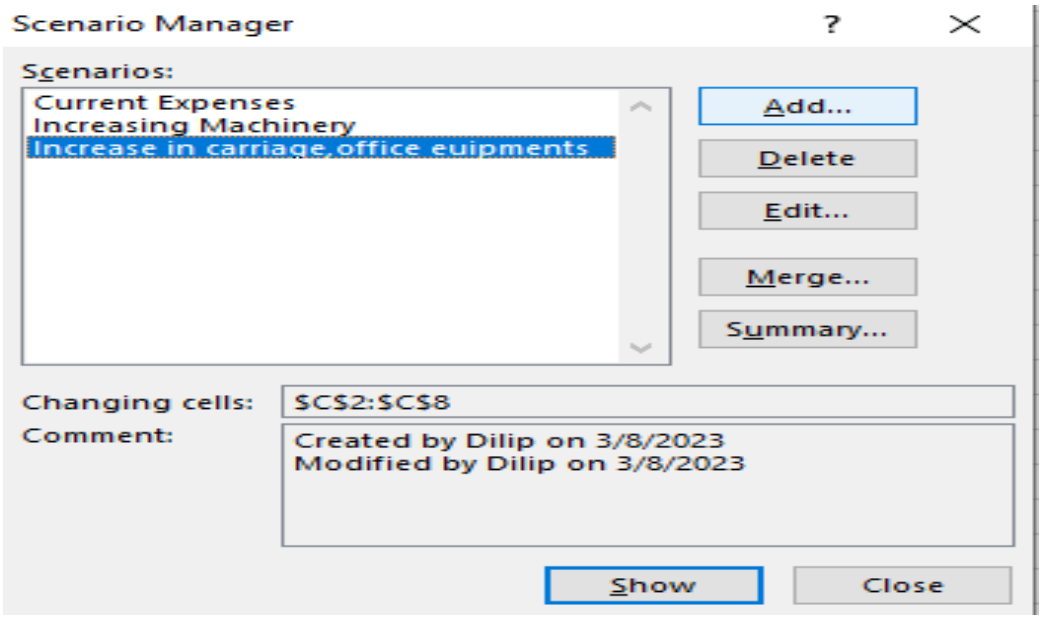
Steps2:In total cost add the formula

SR.NO	Items	Costs
1	Machinery	60000
2	Carriage	8000
3	Transport	30000
4	Office Equipments	6000
5	Postage	7000
6	Miscalleneous	3000
7	Generator	5000
	Total Cost	=SUM(C2:C8)

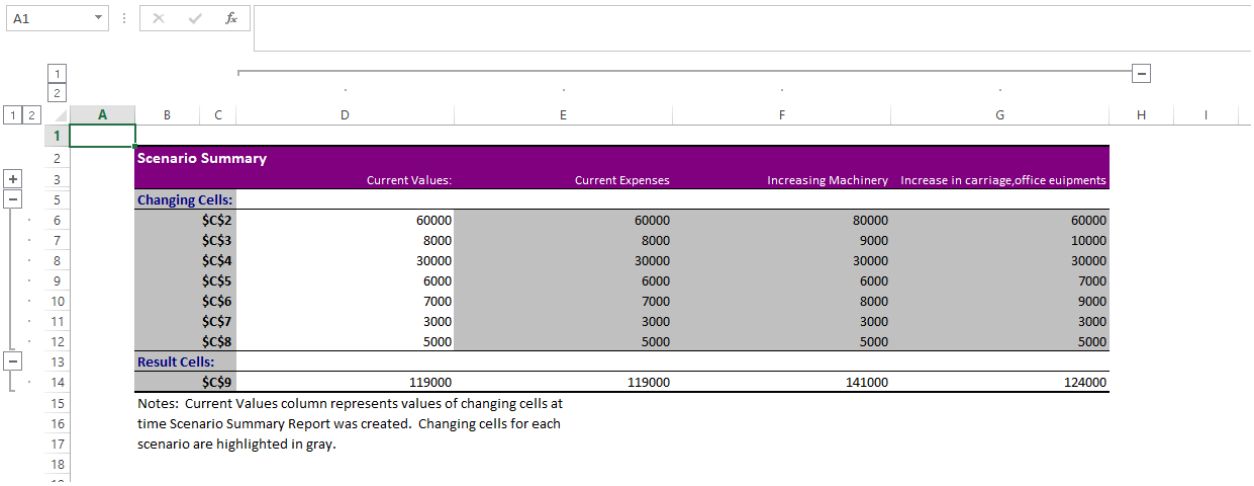
SUM(number1, [number2], ...)

Step3:Go to Data tab→what-if analysis→Scenario manager→add data in it.





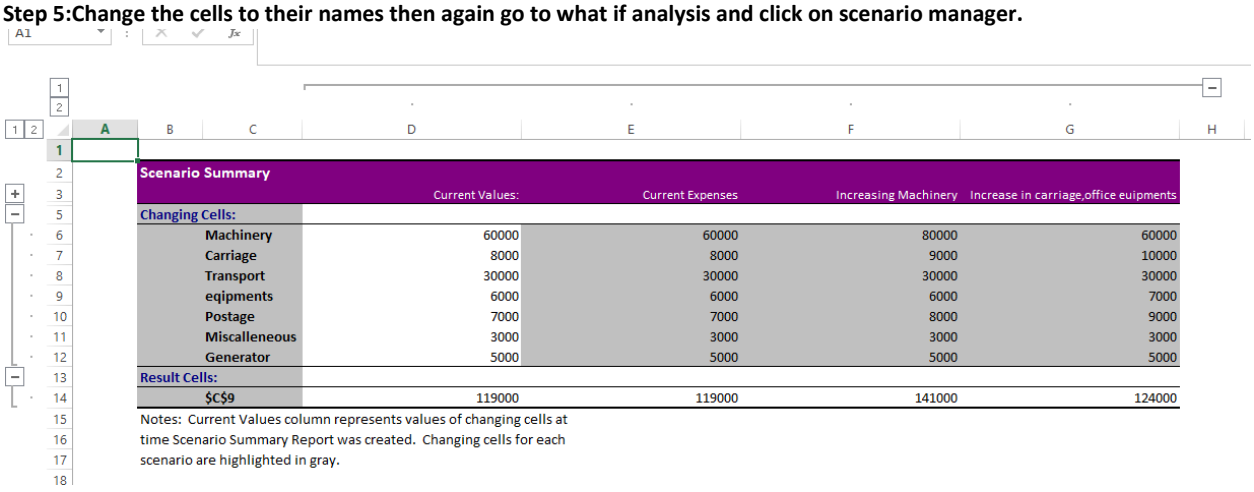
Step 4:After adding scenarios→click on show



Scenario Summary	Current Values:	Current Expenses	Increasing Machinery	Increase in carriage,office euipments
Changing Cells:				
C2	60000	60000	80000	60000
C3	8000	8000	9000	10000
C4	30000	30000	30000	30000
C5	6000	6000	6000	7000
C6	7000	7000	8000	9000
C7	3000	3000	3000	3000
C8	5000	5000	5000	5000
Result Cells:				
C9	119000	119000	141000	124000

Notes: Current Values column represents values of changing cells at time Scenario Summary Report was created. Changing cells for each scenario are highlighted in gray.

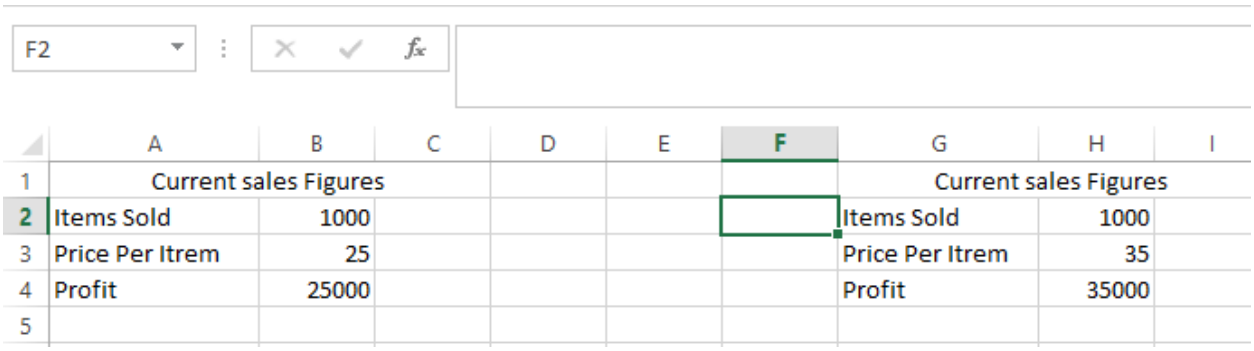
Step 5:Change the cells to their names then again go to what if analysis and click on scenario manager.



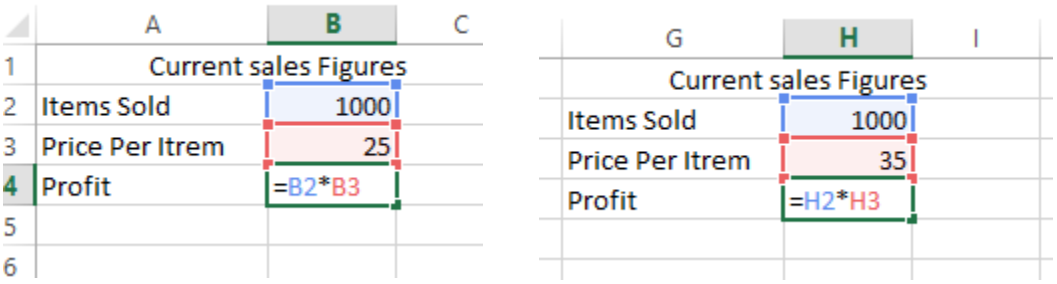
Scenario Summary	Current Values:	Current Expenses	Increasing Machinery	Increase in carriage,office euipments
Changing Cells:				
Machinery	60000	60000	80000	60000
Carriage	8000	8000	9000	10000
Transport	30000	30000	30000	30000
eqipments	6000	6000	6000	7000
Postage	7000	7000	8000	9000
Miscalleneous	3000	3000	3000	3000
Generator	5000	5000	5000	5000
Result Cells:				
C9	119000	119000	141000	124000

Notes: Current Values column represents values of changing cells at time Scenario Summary Report was created. Changing cells for each scenario are highlighted in gray.

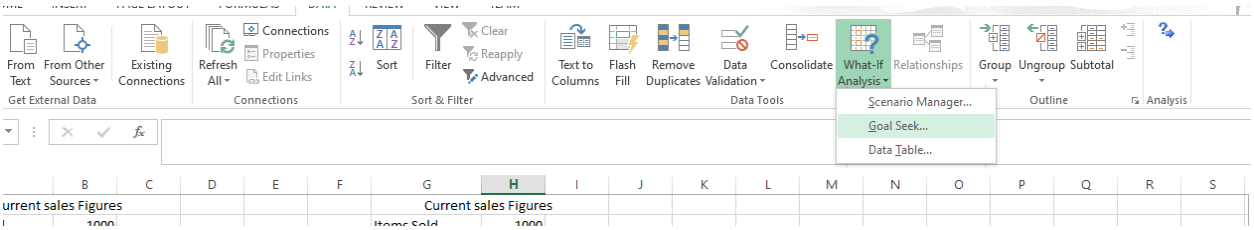
Step 6:Now create a two new tables.



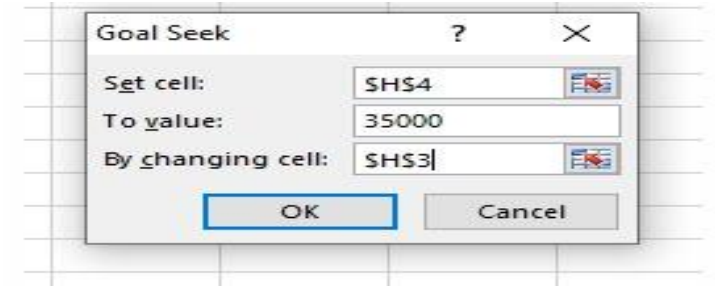
Step 7: Insert formula in profit column of both tables



Step 8:Go to Data tab→what-if analysis→Goal Seek.



Step 9:Set value as→profit column→to value→35000→by changing cell→as price per item



Step 10: Create two tables and insert values in it.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Transportation problem						Shipment	Customer 1	Customer 2	customer 3		Total out	Supply	
2	unit cost	Customer 1	Customer 2	customer 3			Factory 1	0	100	0		100	100	
3	Factory 1	40	47	80			Factory 2	0	100	100		200	200	
4	Factory 2	72	36	58			Factory 3	200	0	100		300	300	
5	Factory 3	24	61	71										
6							Total in	200	200	200				
7													Total cost	
8							Demand	200	200	200			26000	
9														
10														

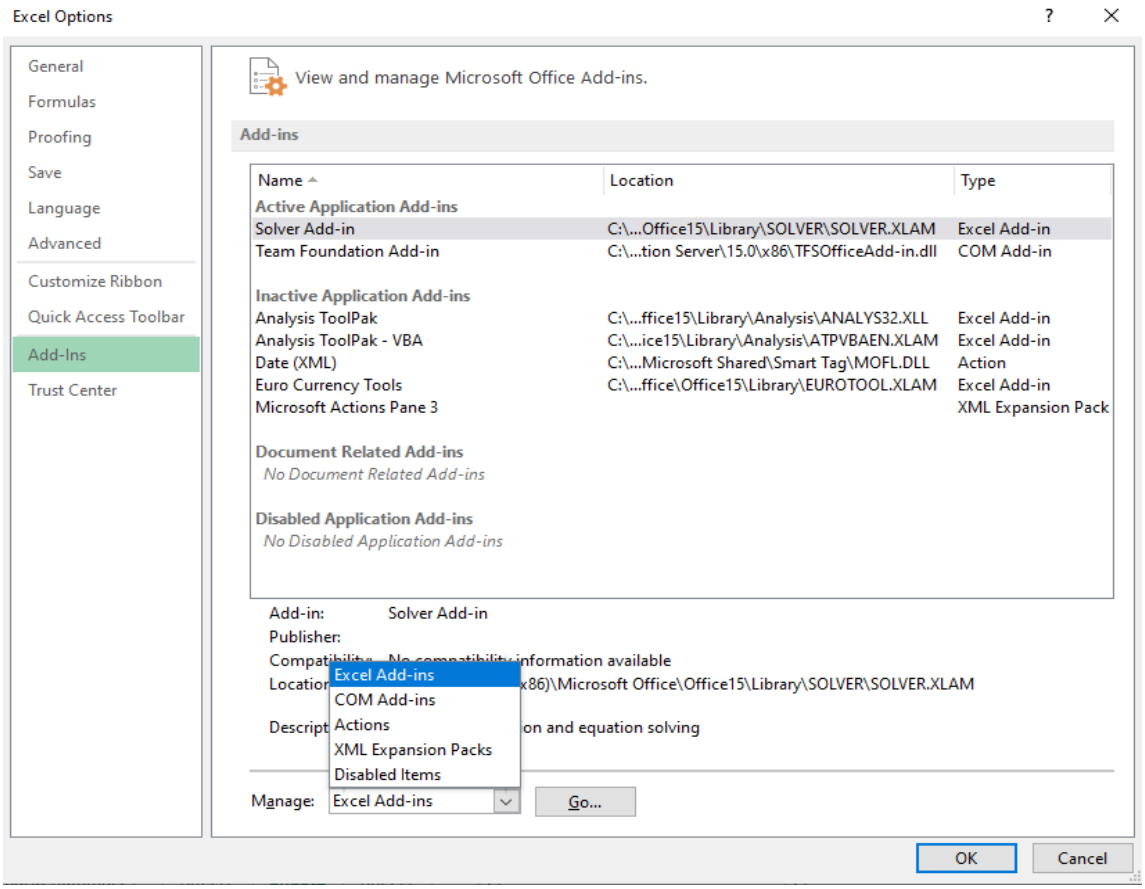
Step 11: Add formulas to Total in, Total out, Total cost.

I	J	K	L	M	N	O	P	Q
stomer 2	customer 3		Total out	Supply				
100	0		=SUM(H2:J2)					
100	100		SUM(**number1**, [number2], ...)					
0	100		300	300				
200	200							
				Total cost				
200	200			26000				

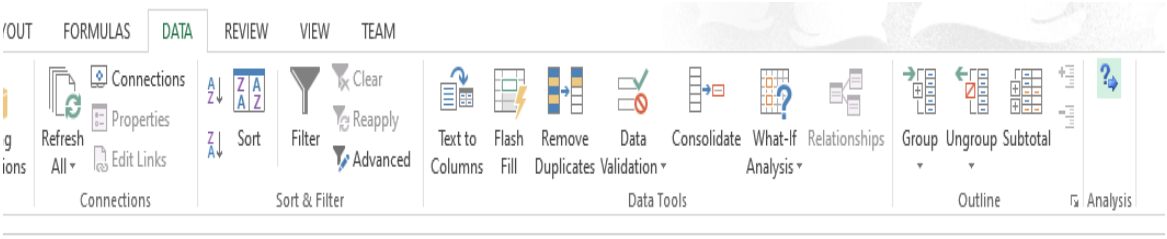
F	G	H	I	J	K	L	M	N
	Shipment	Customer 1	Customer 2	customer 3		Total out	Supply	
	Factory 1	0	100	0		100	100	
	Factory 2	0	100	100		200	200	
	Factory 3	200	0	100		300	300	
	Total in	=SUM(H2:H4)		200				
		SUM(**number1**, [number2], ...)					Total cost	
	Demand	200	200	200			26000	

G	H	I	J	K	L	M	N	O	P	Q
Shipment	Customer 1	Customer 2	customer 3		Total out	Supply				
Factory 1	0	100	0		100	100				
Factory 2	0	100	100		200	200				
Factory 3	200	0	100		300	300				
Total in	200	200	200							
						Total cost				
Demand	200	200	200			=H2*B3+I2*C3+J2*D3+H3*B4+I3*C4+J3*D4+H4*B5+I4*C5+J4*D5				

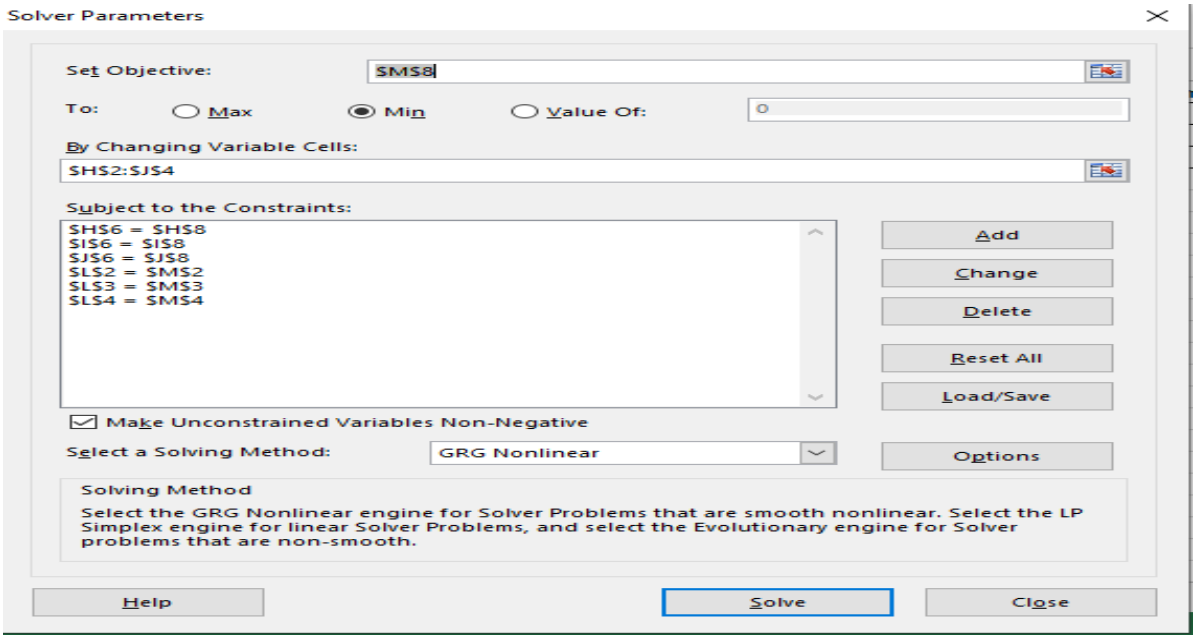
Step 12:Go to file →options→add-ins→excel add-ins→go→ok.



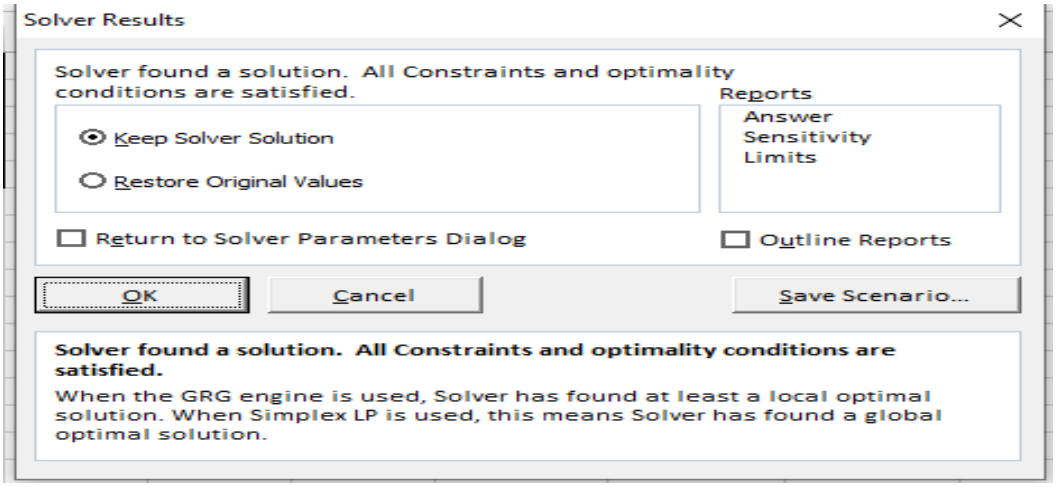
Step 13:Go to data tab→analysis→Solver



Step 14:Click on solver and add set objective as the total cost cell→changing cells→select all the columns of shipment table→constraints→cells of demand and supply→click on solve.



Step 15:After clicking on solve→click ok.



Step 16:All values are automatically inserted in shipment table. With its total cost.

F	G	H	I	J	K	L	M	N
	Shipment	Customer 1	Customer 2	customer 3		Total out	Supply	
	Factory 1	0	100	0		100	100	
	Factory 2	0	100	100		200	200	
	Factory 3	200	0	100		300	300	
	Total in	200	200	200				
							Total cost	
	Demand	200	200	200			26000	