

Data Mining: Clustering



Applies to:

SAP BI 7.0. For more information, visit the [EDW homepage](#)

Summary

This article deals with Data Mining and it explains the classification method 'Clustering' in detail. It also explains the steps for implementation of Clustering by creating a Model and an Analysis Process.

Author: Vishall Pradeep K.S

Company: Applexus Technologies (P) Ltd

Created on: 1 May 2011

Author Bio

Vishall Pradeep is working as SAP Technology Consultant with Applexus Technologies (P) Ltd. He has experience in SAP ABAP and SAP BI

Table of Contents

Introduction	3
Clustering	3
Creating a Model.....	3
Creating a Analysis Process for Training.....	6
Creating a Analysis Process for Executing a Prediction	13
Related Content.....	21
Disclaimer and Liability Notice.....	22

Introduction

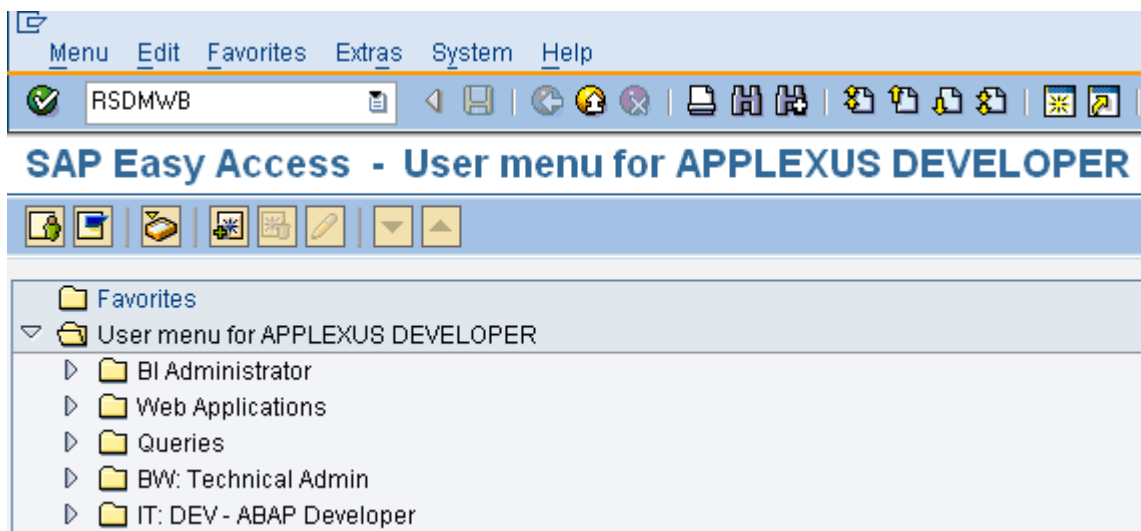
Data mining is to automatically determine significant patterns and hidden associations from large amounts of data. Data mining provides you with insights and correlations that had formerly gone unrecognized or been ignored because it had not been considered possible to analyze them. The data mining methods available in SAP BW allow you to create models according to your requirements and then use these models to draw information from your SAP BW data to assist your decision-making.

Clustering

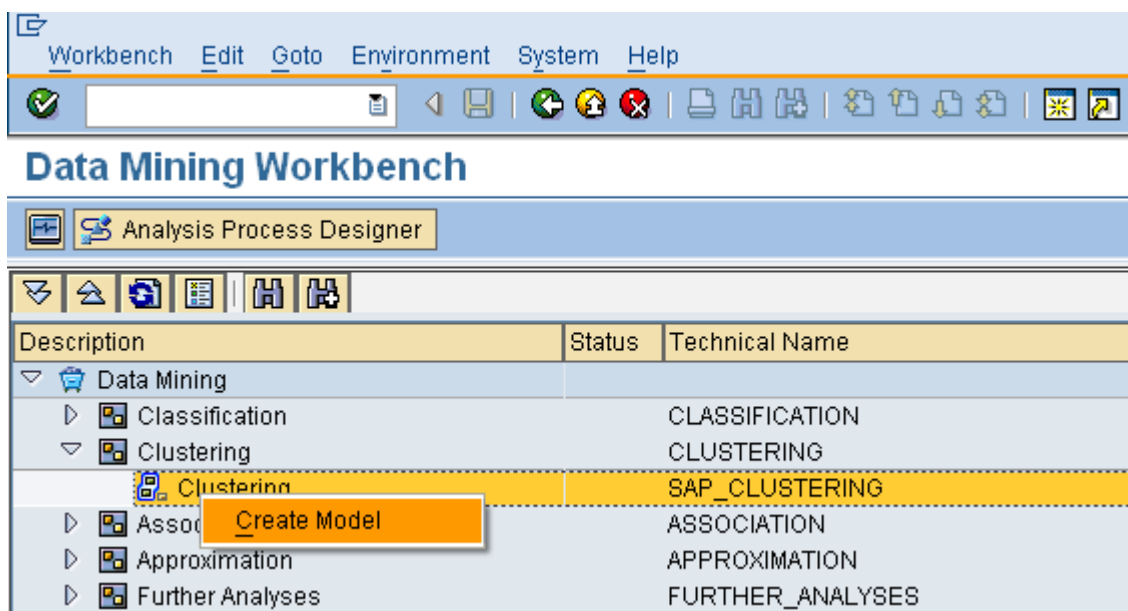
Clustering allows you to segment data automatically into clusters. In a subordinate dataset, the system groups together associated data by forging formerly unknown links. This entails determining the criteria for clustering as well as the mappings between datasets

Creating a Model

- Go to Transaction RSDMWB (Data Mining Workbench)



- Data Mining->Expand Clustering->Right Click Clustering->Create Model



- Choose the Model Name and Description
- The method name for which you are creating a model is displayed. You have three options for model field selection

The screenshot shows the 'Create Model' dialog box with the following fields and options:

- Model Name: ZPDT_MODEL_CLUST
- Description: Model for Clustering
- Method: (empty)
- Options:
 - Manual
 - Use Model as Template
 - Name: (empty)
 - Model Field Selection
 - BW Query: (empty)

- To create the model fields manually, select the Manual option.
- If you want to create a model that is similar to an existing model created previously, you can copy it choosing the Use Model as Template option. You can make minor changes to the copied version manually to suit your requirements
- To create a model from a query, choose Model Field Selection and select the query which you want use as a source for model fields .The InfoObjects contained in the selected query are available as model fields

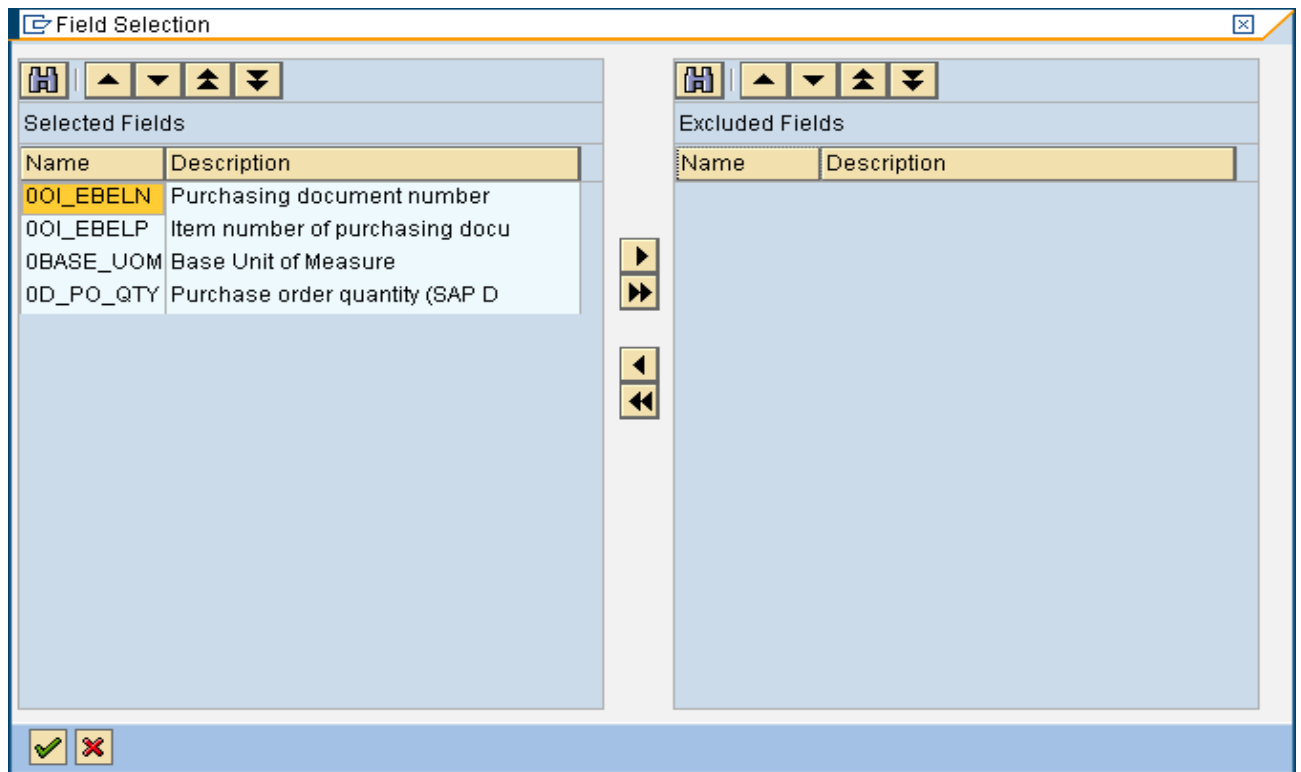
The screenshot shows the 'Queries and Views' interface with the following elements:

- Menu: Queries, Edit, Goto, System, Help
- Toolbar: Refresh, Technical Names, and various navigation icons.
- Available Objects:
 - IC for Ekpo
 - Purchase Order Test - Vishall (selected)

The screenshot shows the 'Create Model' dialog box with the following fields and options:

- Model Name: ZPDT_MODEL_CLUST
- Description: Model for Clustering
- Method: (empty)
- Options:
 - Manual
 - Use Model as Template
 - Name: (empty)
 - Model Field Selection
 - BW Query: I_IC_VEKPO/REP_20110420063408

- The screen shows the list of Fields and we can select and exclude fields in it



- In the *Fields* tab, to specify which characteristic is to be considered with which attributes and field parameters are used to specify weightings for the individual attributes. The system then establishes formerly unknown associations between the attribute values

Name	Description	InfoObject	Data ...	Leng..	Content Type	Paramet..	Values
0BASE_UOM	Base Unit of Measure	0BASE_UOM	UNIT	3	Discrete		123
0D_PO_QTY	Purchase order quantity (SAP D	0D_PO_QTY	QUAN	23	Continuous		123
00I_EBELN	Purchasing document number	00I_EBELN	CHAR	10	Discrete		123
00I_EBELP	Item number of purchasing docu	00I_EBELP	NUMC	5	Continuous		123

- The *Content types* valid for a model field are dependent on the method that you are creating the model for and on the data type of the model field. The value type specified for a model field determines which entries can be made as Field Parameters and Field Values.
- In the *Parameters* tab, we can specify how many clusters the system should create during training and by specifying conditions for interrupting the segmentation; you enhance the quality and performance of the segmentation.

What If Export as PMML

Header

Name	ZPDT_MODEL_CLUST	Description	Model for
Version	Revised	State	◇
Created by		Modified by	
Created on		Modified on	
Method			

Fields Parameters

Model Parameters

No. of Clusters	5
Max. Distinct Values allowed	25

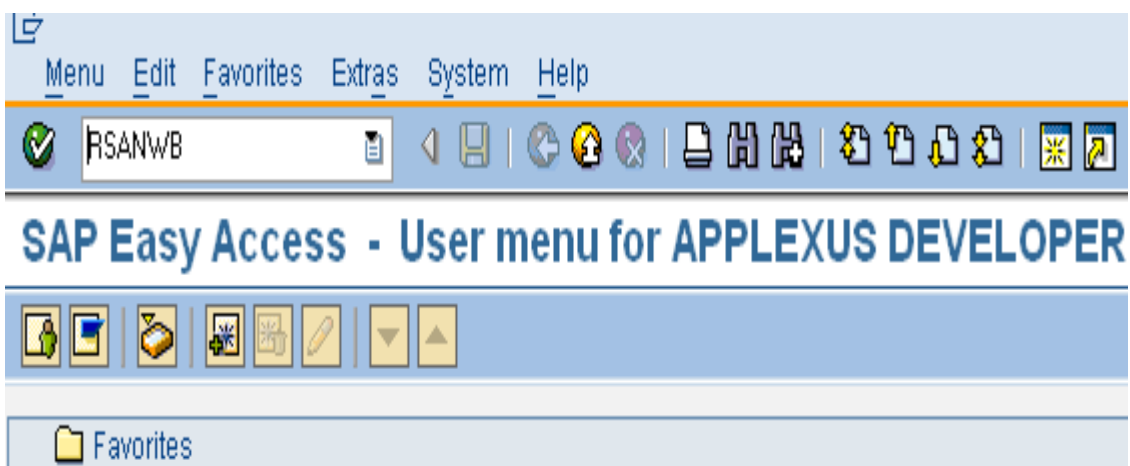
Stopping Criteria

Max. no. of Iterations	5
Min. Fraction of Inter Cluster hops	,001

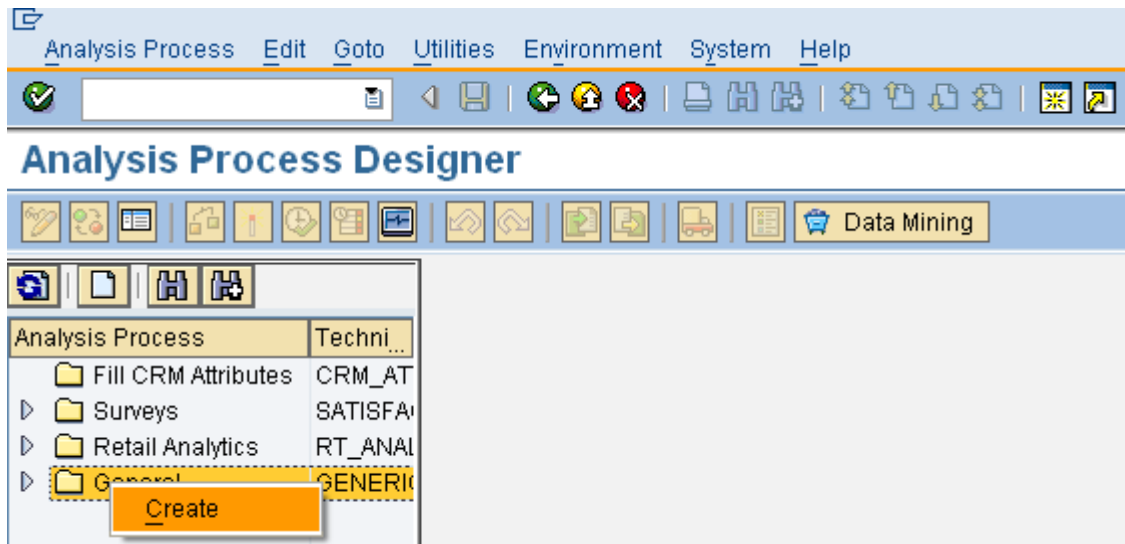
- Save and Activate the Model (we can only train or evaluate a model or use it for the prediction if the model has been activated.)

Creating a Analysis Process for Training

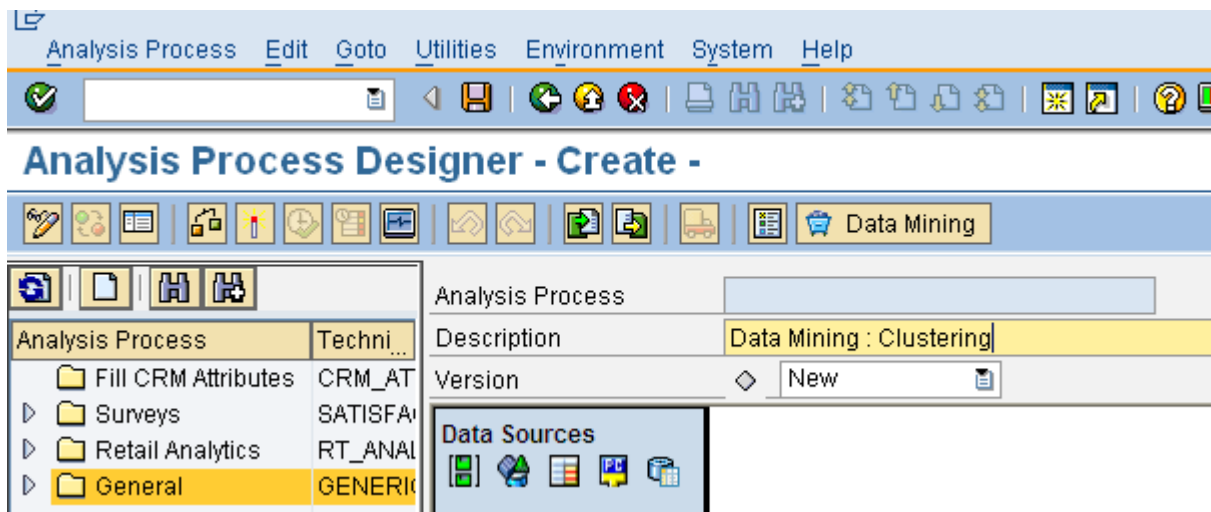
- Go to Transaction RSANWB (Analysis Process Designer)



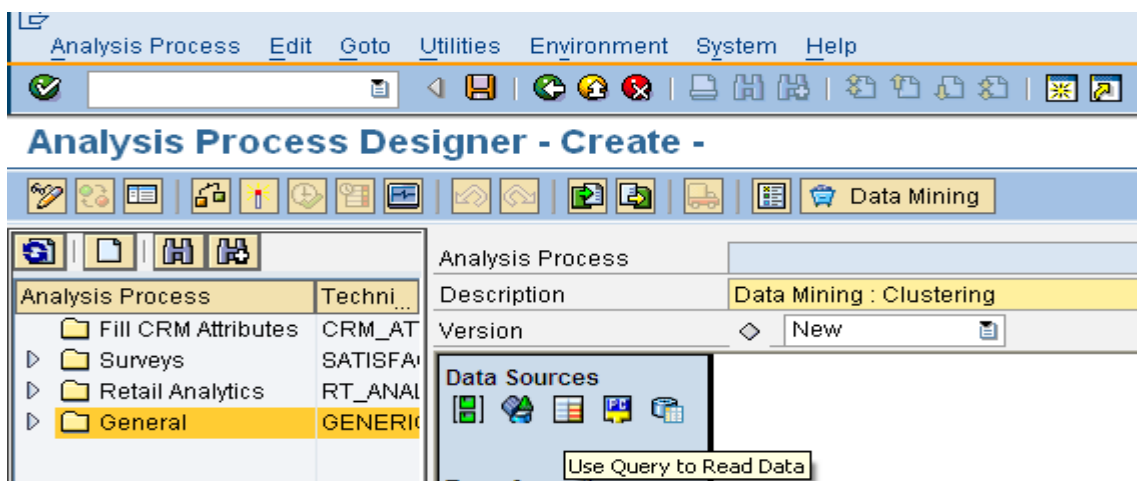
- Choose General->Right Click->Create



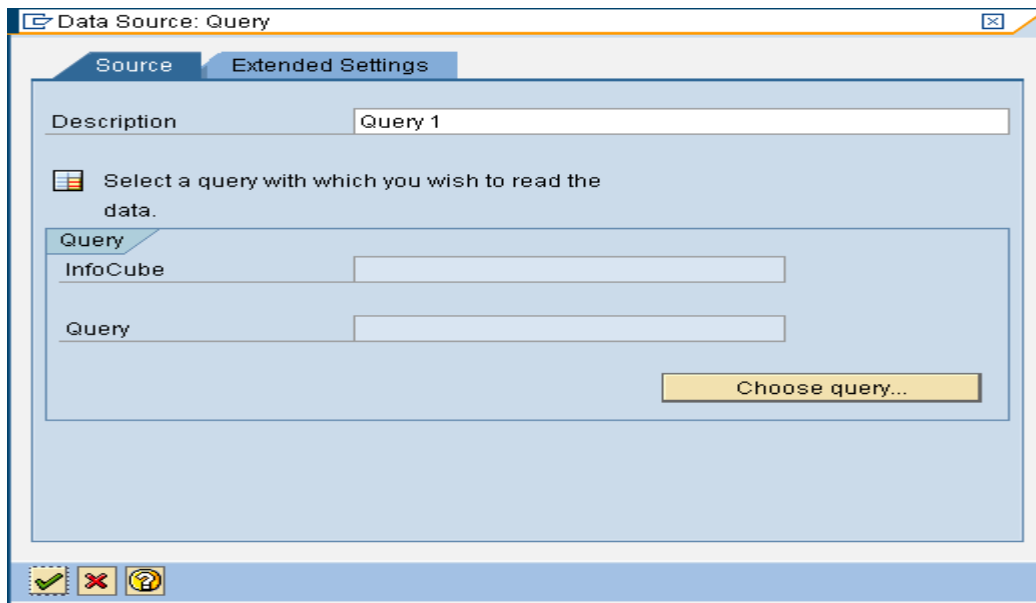
- Give the description to the APD



- From the Data Sources, drag and drop the Query to the work area



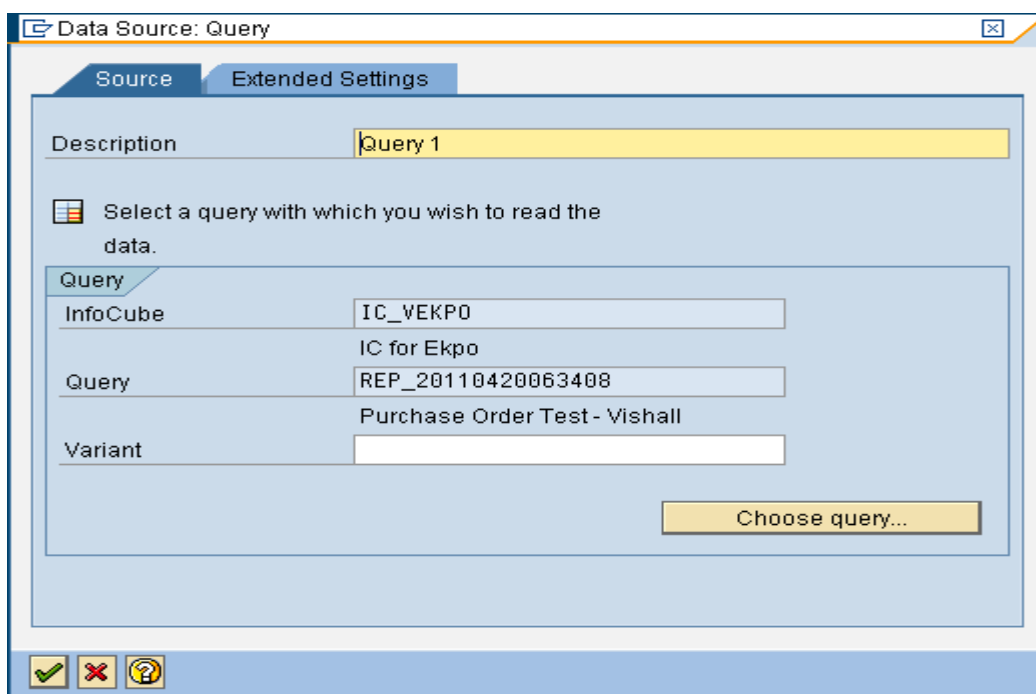
- It asks for a Popup and click on Choose Query



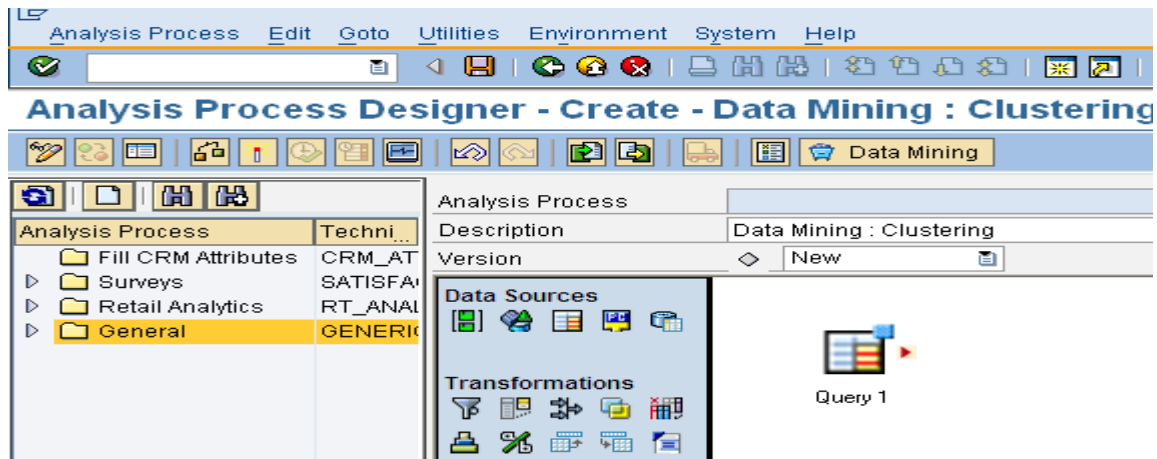
- From the Help, Select the query



- And Click "OK"



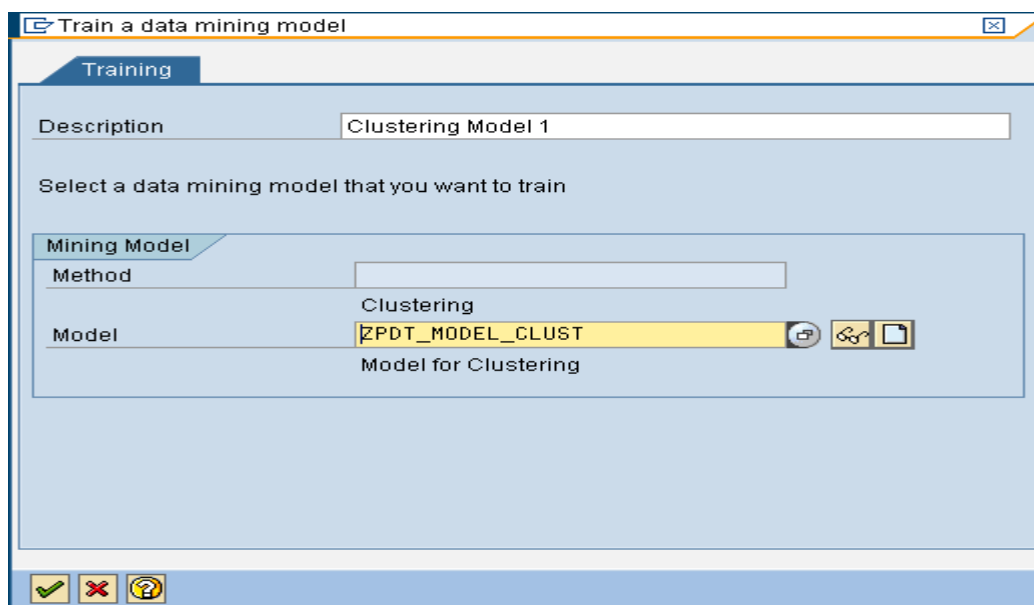
- The Query which is the data Source is added as below



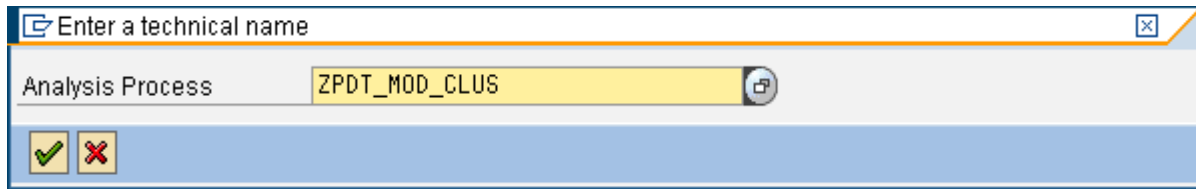
- For the data target, drag the icon for the relevant data mining method in the work area



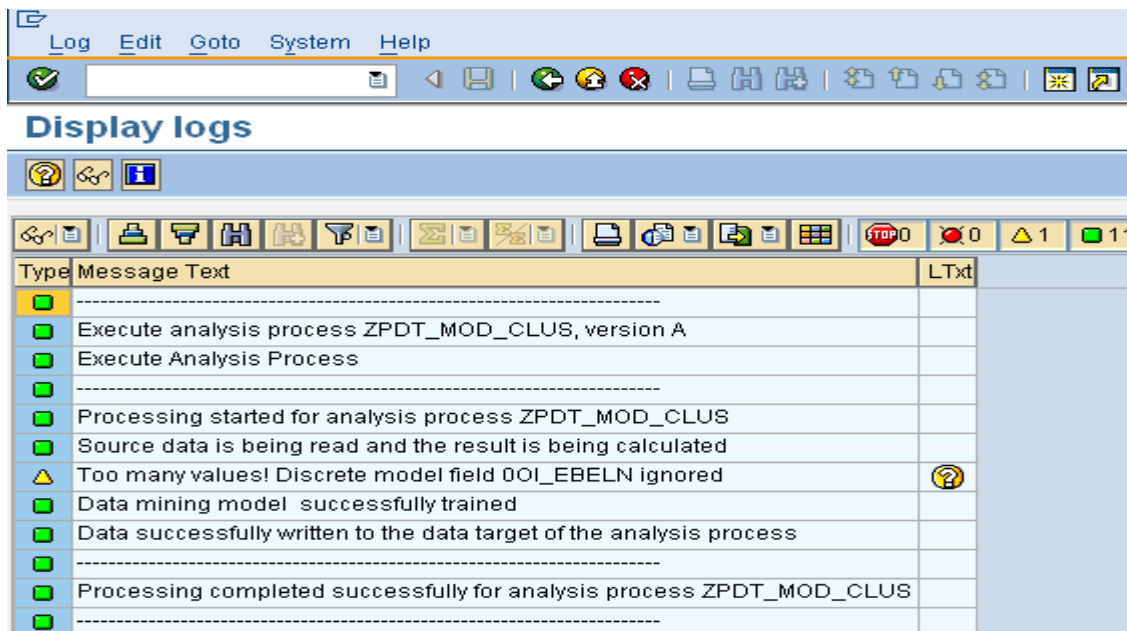
- Double click on data mining node to make the settings in the dialog box that appears
- Choose the required model from F4 Help



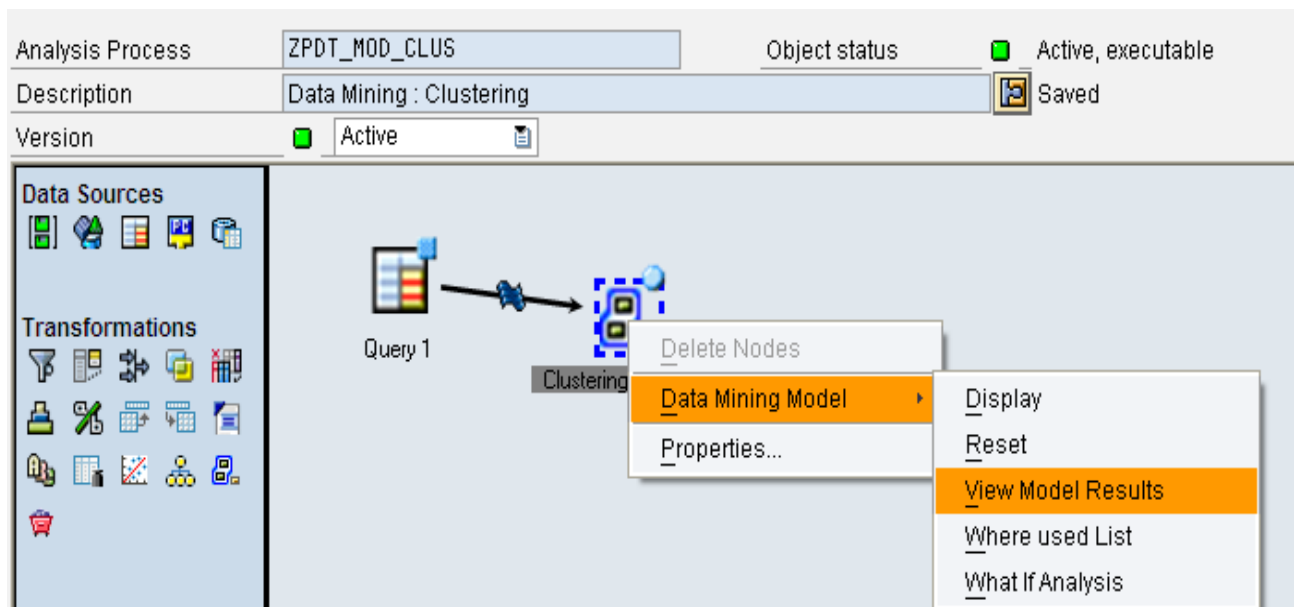
- Click on Continue and Save and activate the APD
- While saving it will ask for a Technical Name



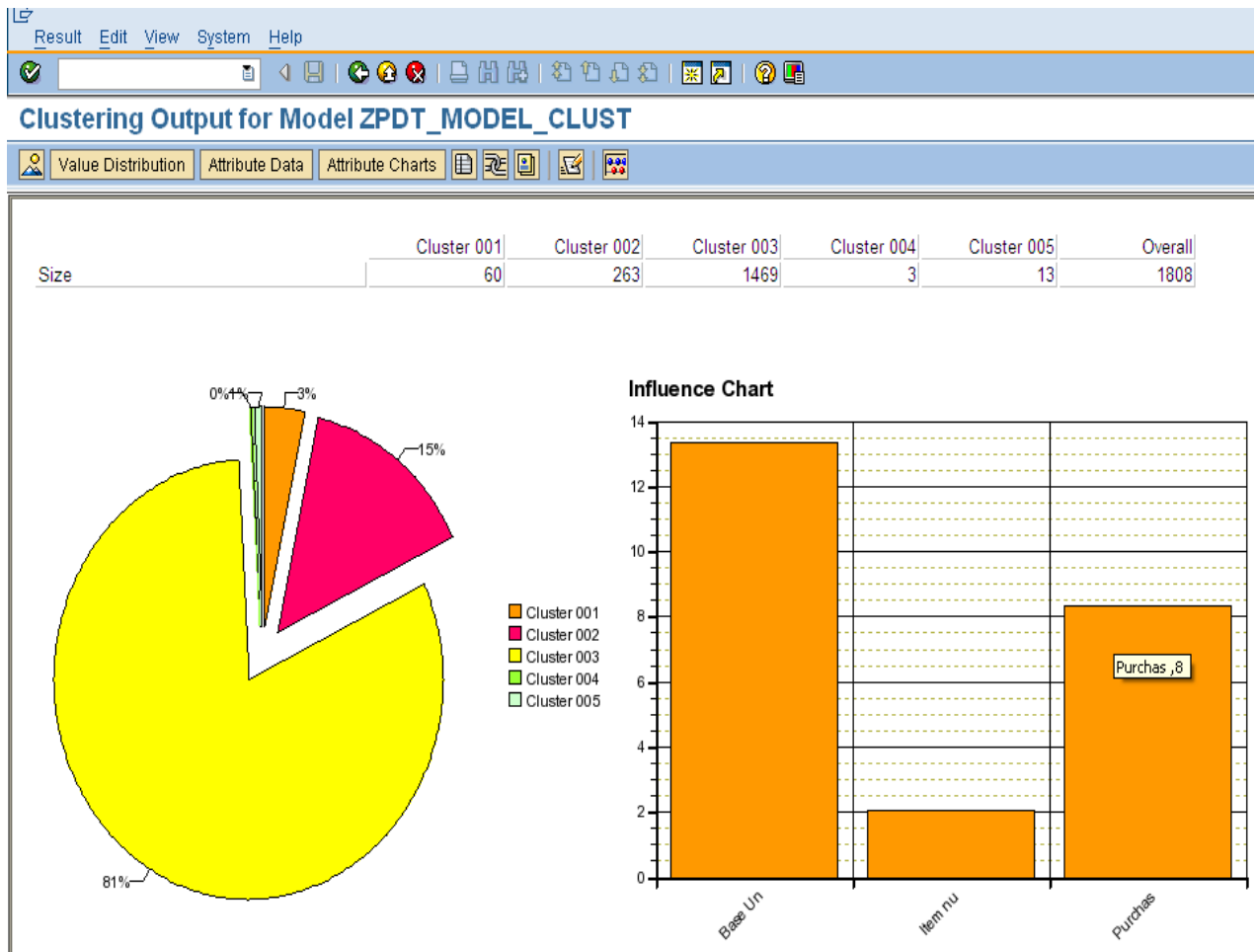
- Execute the APD
- The data is written to the data target and a log is displayed



- To view the training results, in the context menu of data target, choose *Data Mining Model* → *View Model Results*



- The Results will be shown as below

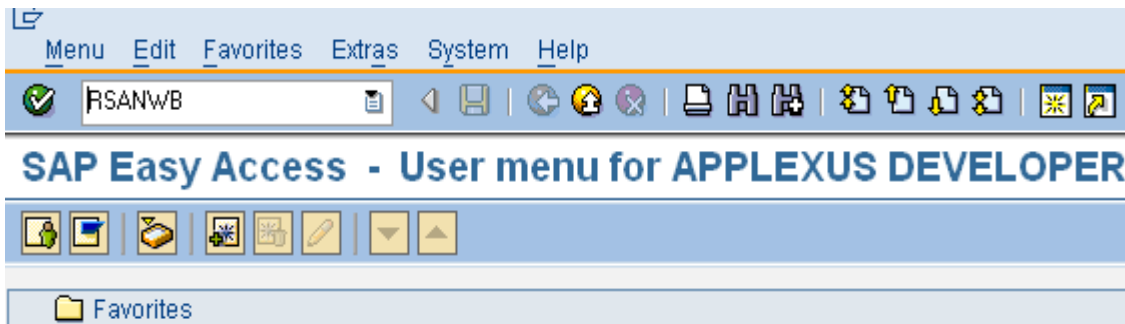


- By Using Value Distribution button you can “ View Attribute Value Distribution as Graphic” and Attribute Data to “ View Attribute Value Distribution as Tables” and Attribute Charts to “ View Attribute Value Distribution across Clusters”
- The data mining model acquires the status *Trained*

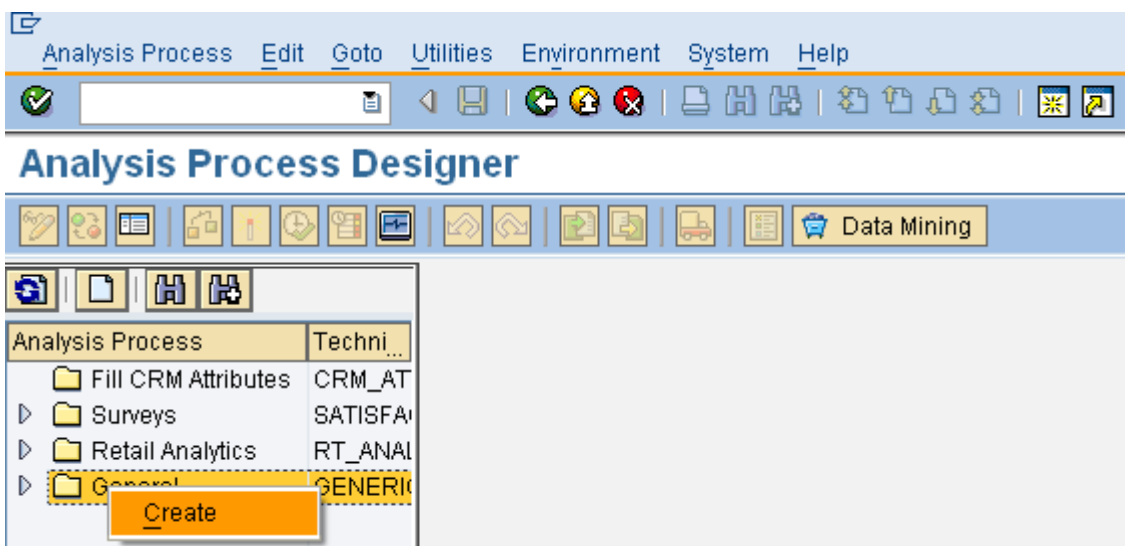
Header			
Name	ZPDT_MODEL_CLUST	Description	Model for Clustering
Version	Active	State	OO
Created by	APPDEVELOP	Modified by	APPDEVELOP
Created on	04/29/2011	Modified on	04/29/2011

Creating a Analysis Process for Executing a Prediction

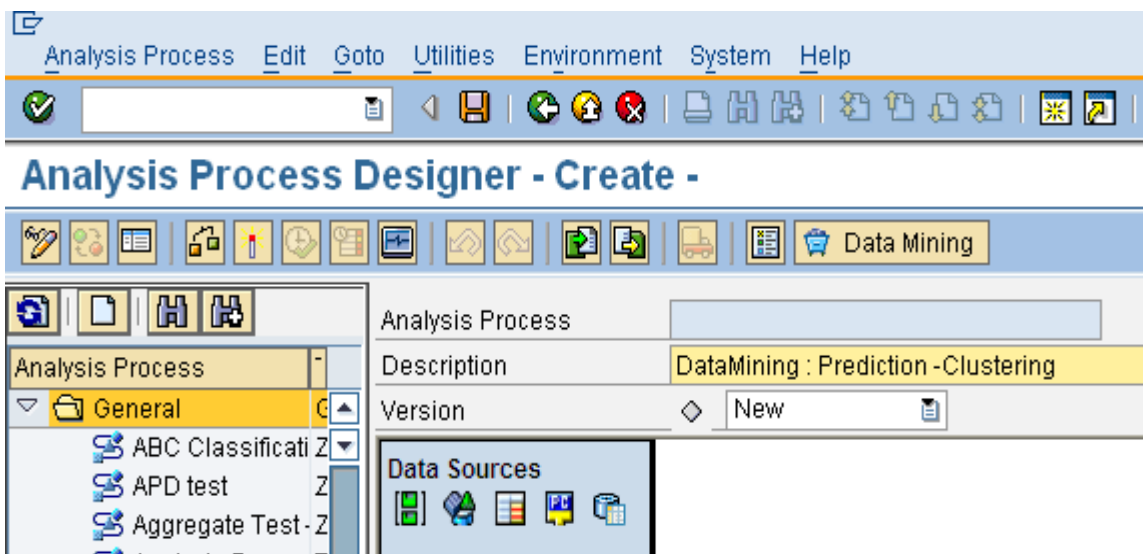
- A model that you trained using historic data from a source can now be applied to a different set of data and the predicted output could be the best of three clusters
- Goto Transaction RSANWB (Analysis Process Designer)



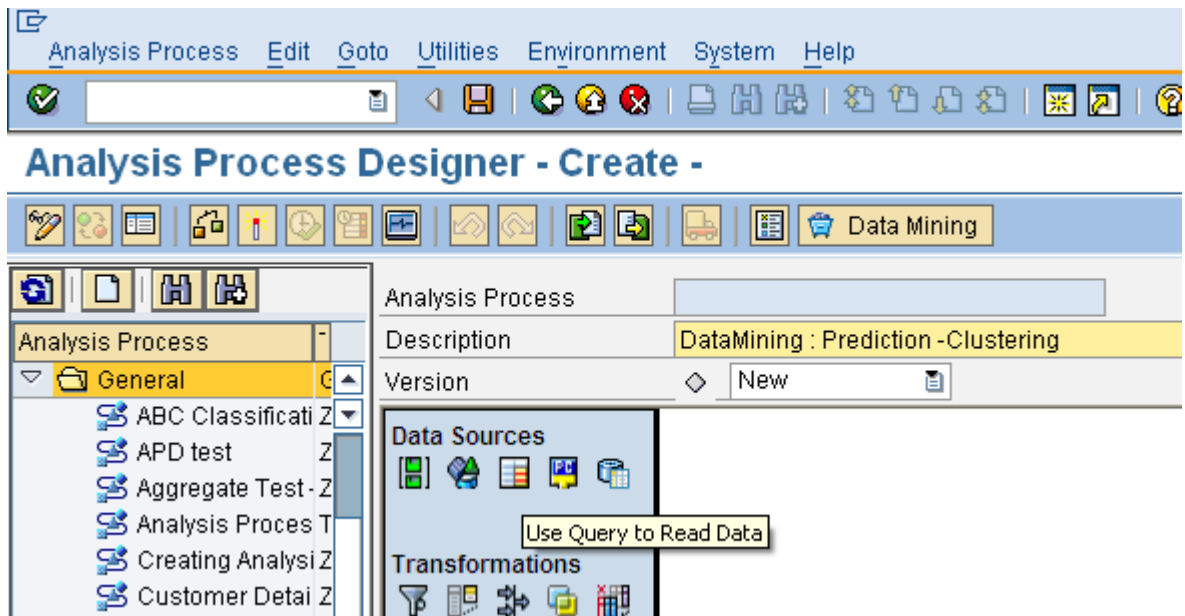
- Choose General->Right Click->Create



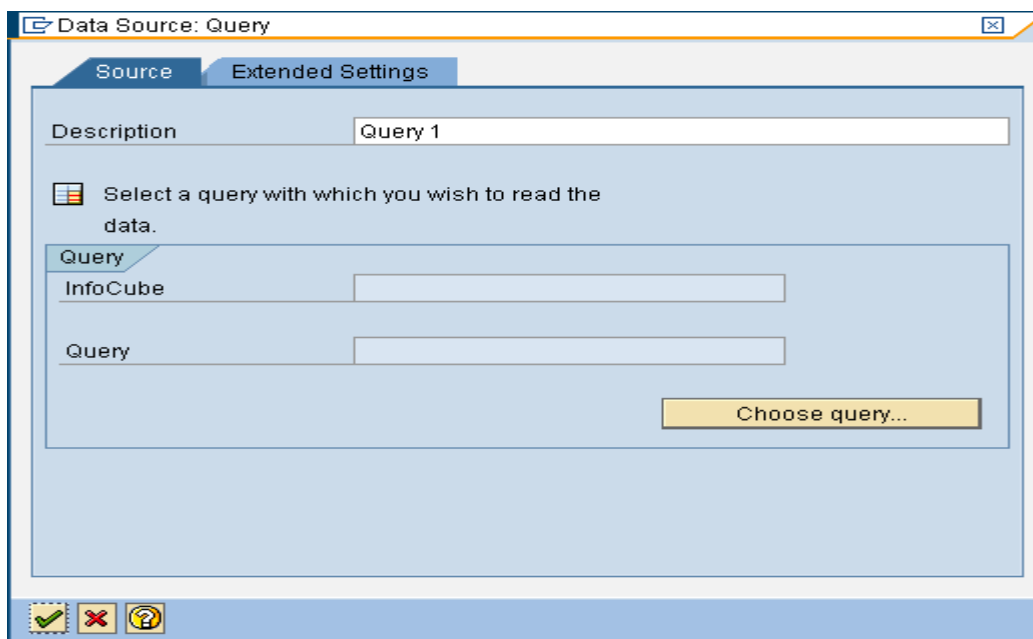
- Give the description to the APD



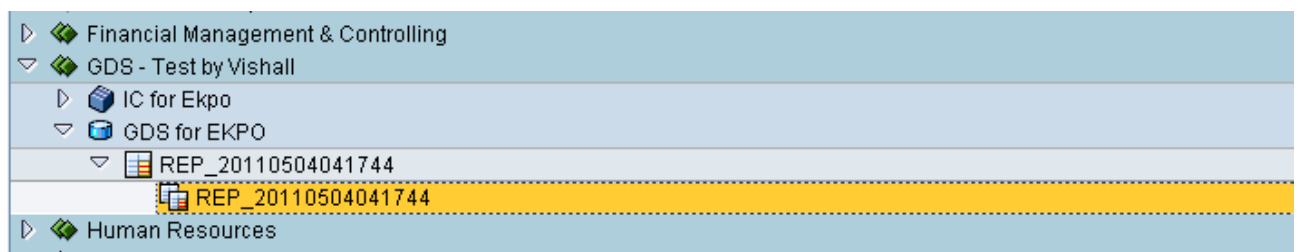
- From the Data Sources, drag and drop the Query to the work area



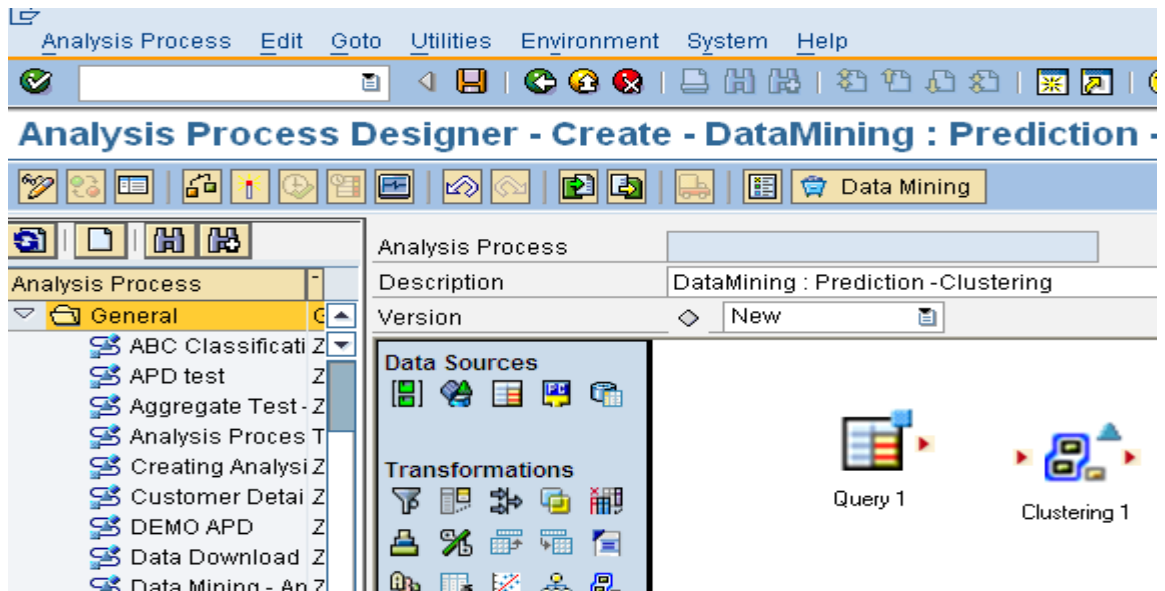
- It asks for a Popup and click on Choose Query



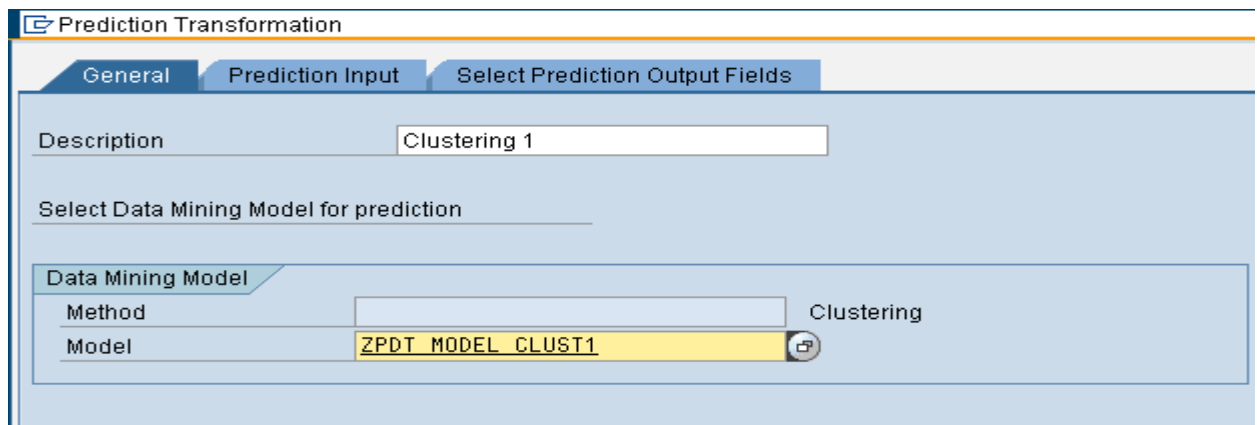
- From the Help, Select the query



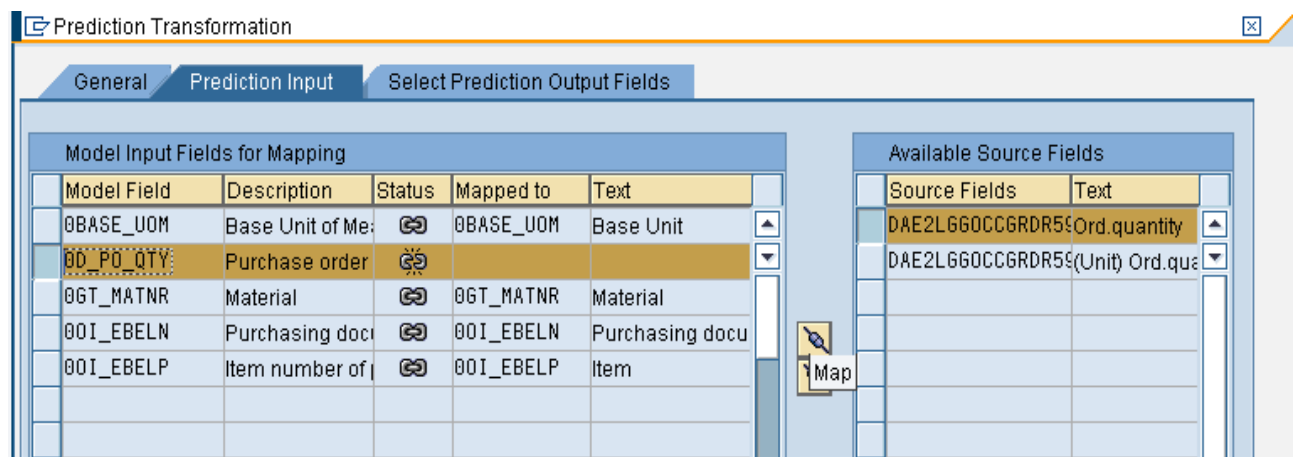
- Drag the relevant prediction icon, that is, source for transformation, in the work area



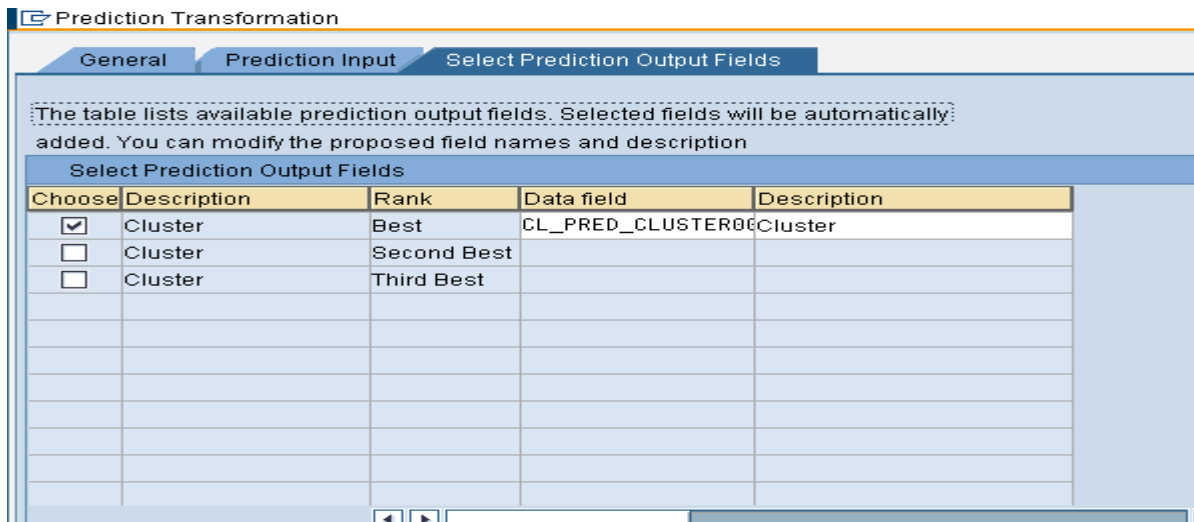
- Connect the two nodes
- Double click on data mining node to make the settings in the dialog box that appears
- Choose the required model from F4 Help



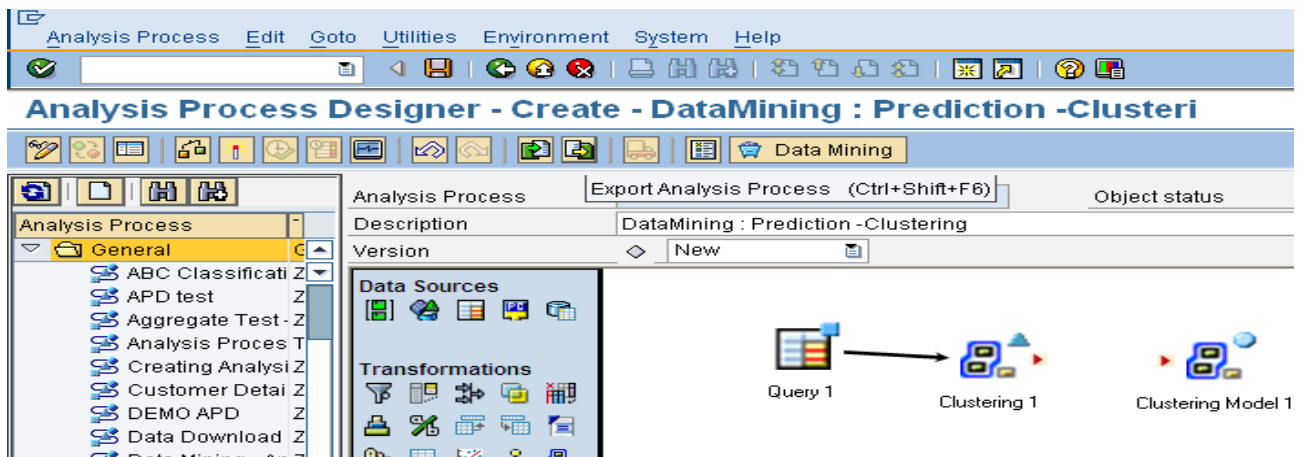
- In Prediction Input Tab , Map the Available source field and Input field for prediction



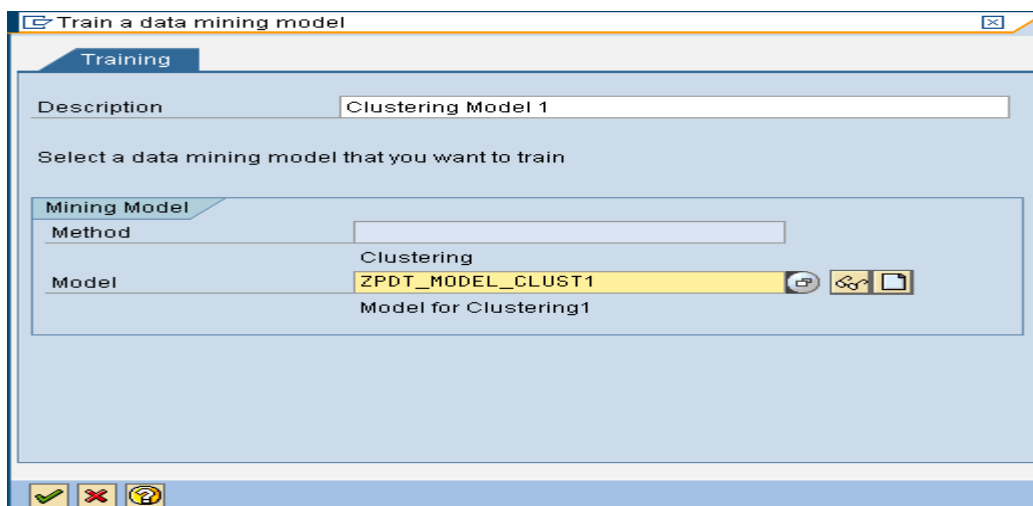
- In select Prediction Output Fields “Select the Fields”



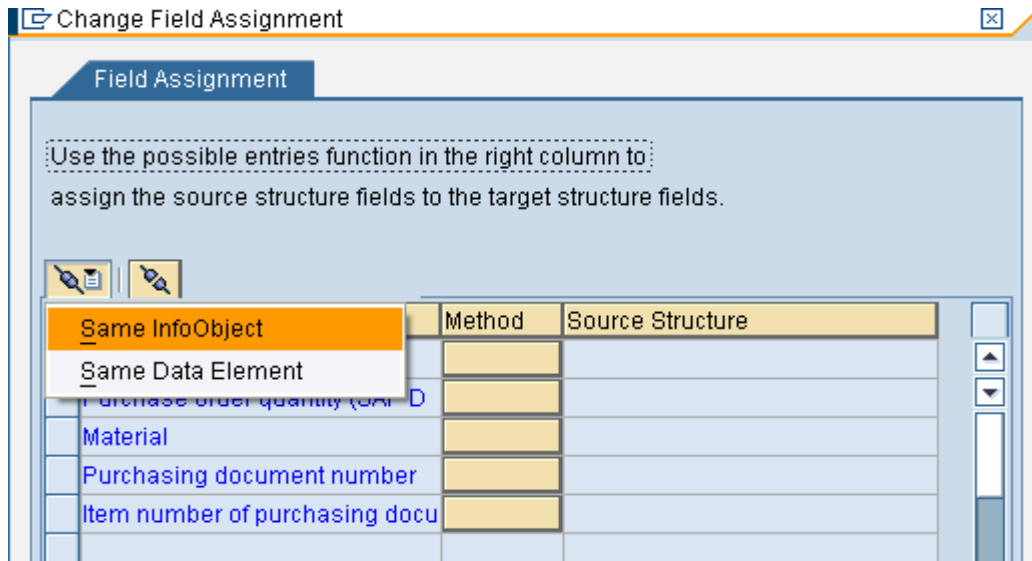
- For the data target, drag the icon for the relevant data mining method in the work area (In this case I am downloading the predicted values to another Cluster Model)
- Create a cluster model in “RSDMWB” and Keep it in New State (Don’t Train the Model)



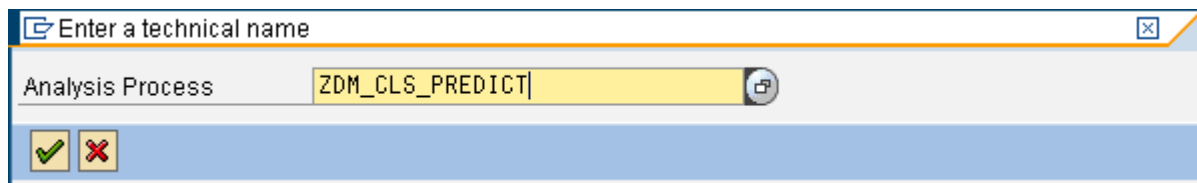
- Double click the Target and Choose the model



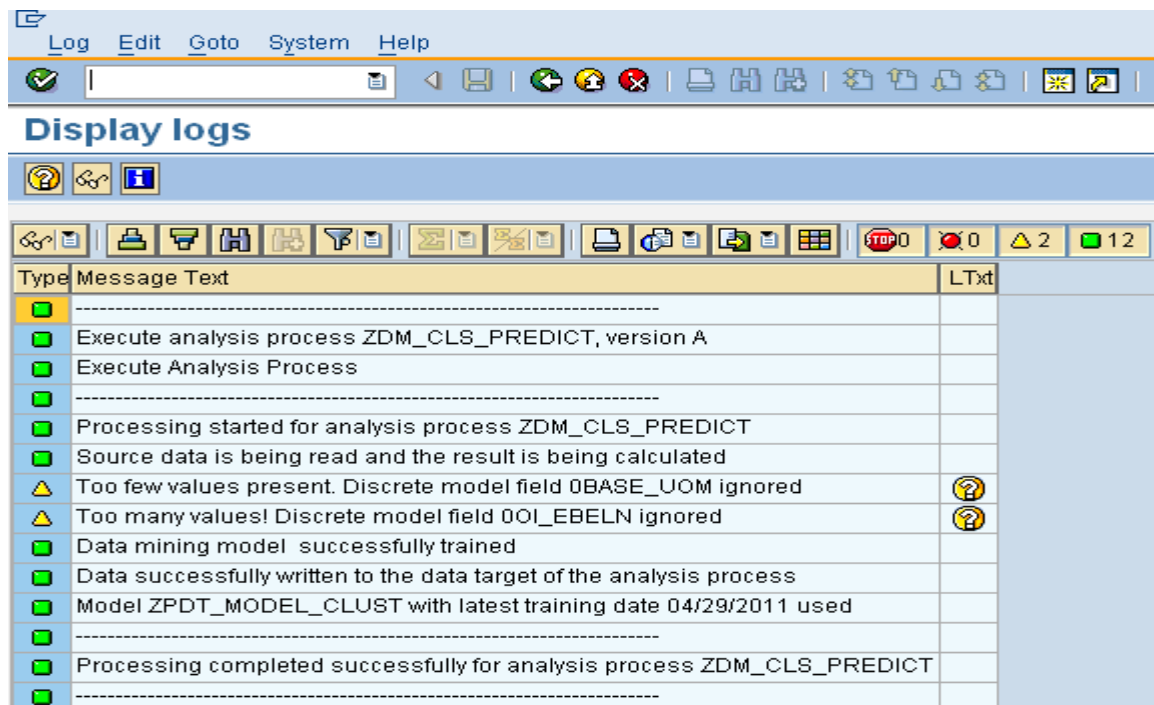
- Connect the nodes
- Double Click on the Node Connection and assign the source and target structure fields



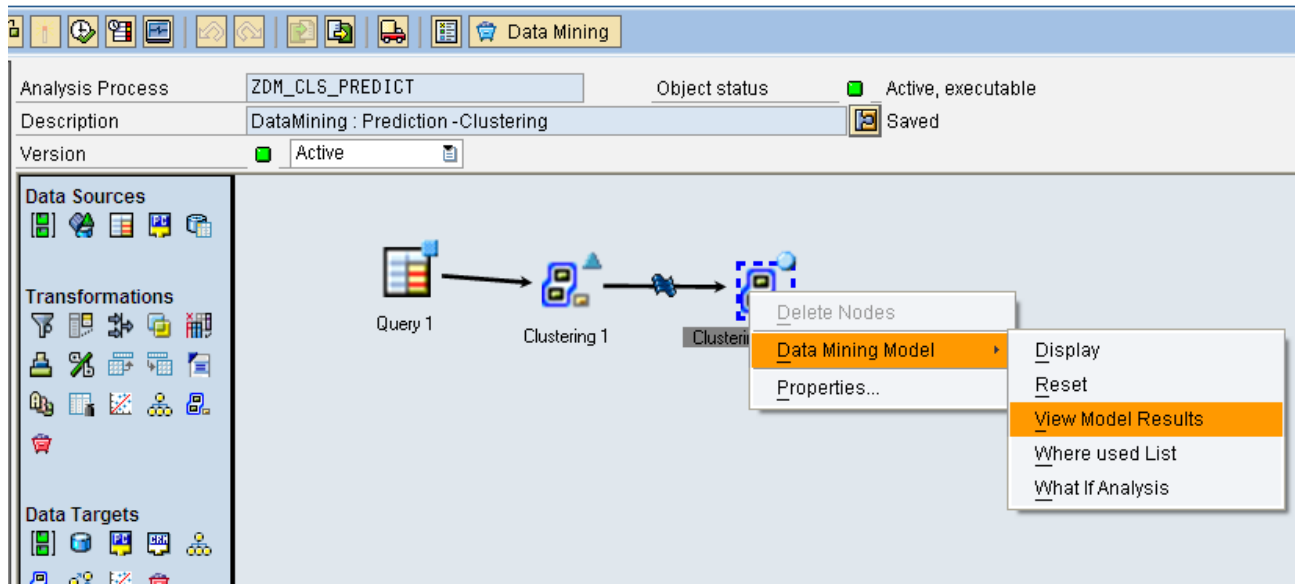
- Save and activate the APD
- While saving it will ask for a Technical Name



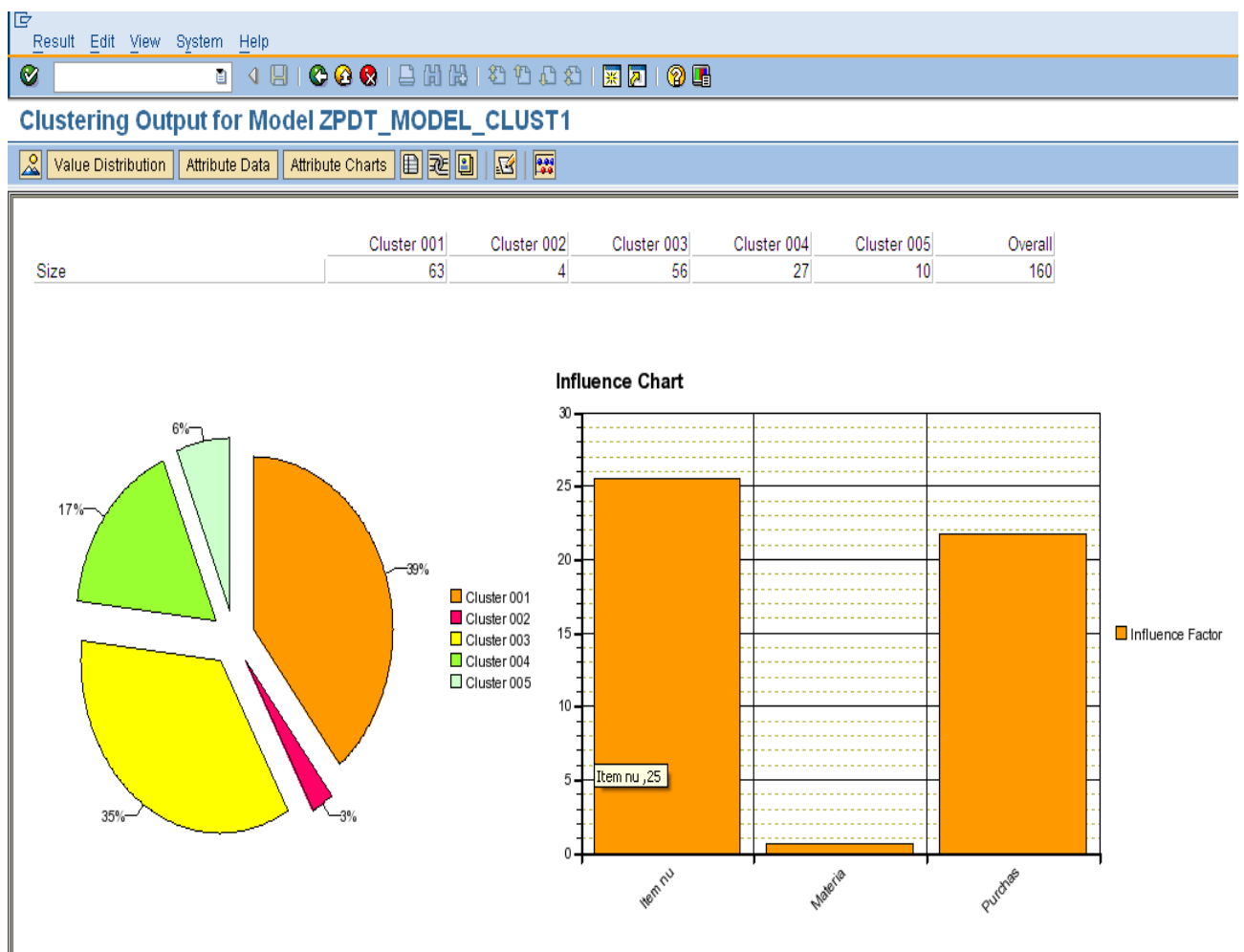
- Execute the APD



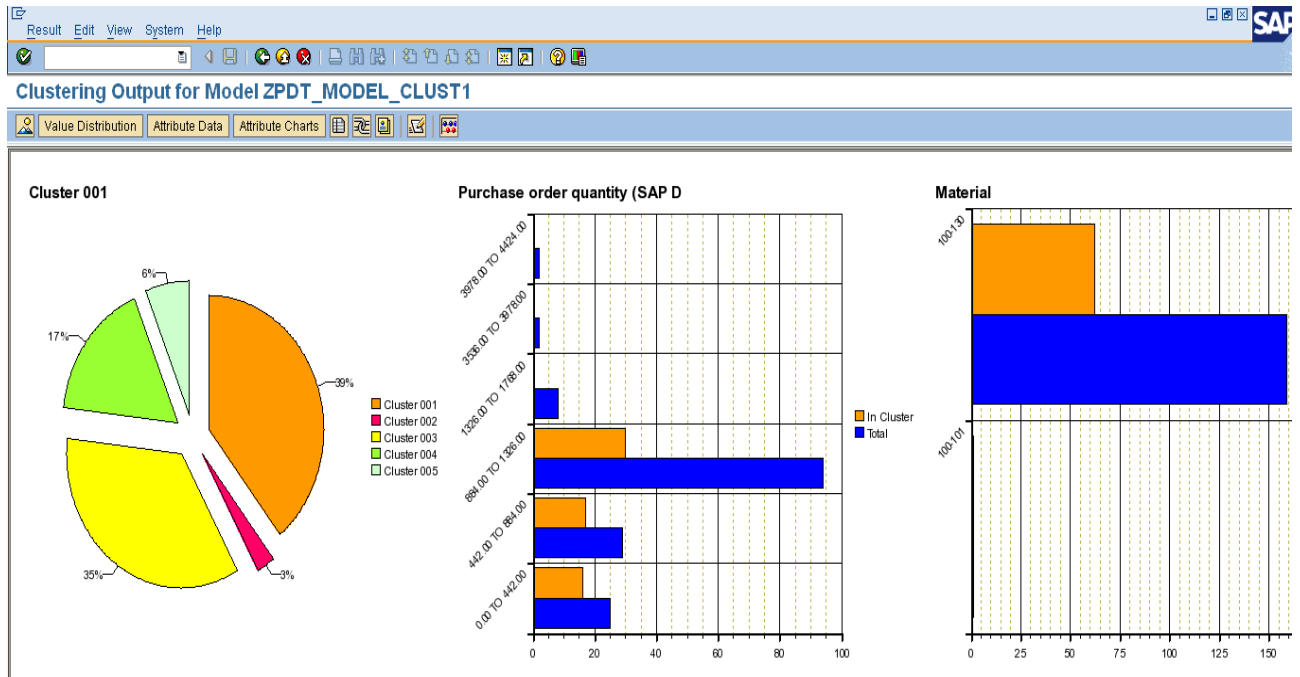
- To view the Prediction results, in the context menu of data target, choose *Data Mining Model* → *View Model Results*



- The Results will be shown as below



- Click on Distribution button you can “ View Attribute Value Distribution as Graphic”



- Click on Attribute Data to “ View Attribute Value Distribution as Tables”

The screenshot shows the SAP Clustering Output for Model ZPDT_MODEL_CLUST1, displaying three tables under the 'Attribute Data' tab:

Purchase order quantity (SAP D)

	Cluster 001	Cluster 002	Cluster 003	Cluster 004	Cluster 005	Overall
0.00 TO 442.00	16	0	0	0	9	25
442.00 TO 884.00	17	0	0	11	1	29
884.00 TO 1326.00	30	0	48	16	0	94
1326.00 TO 1768.00	0	0	8	0	0	8
1768.00 TO 2210.00	0	2	0	0	0	2
2210.00 TO 2652.00	0	2	0	0	0	2

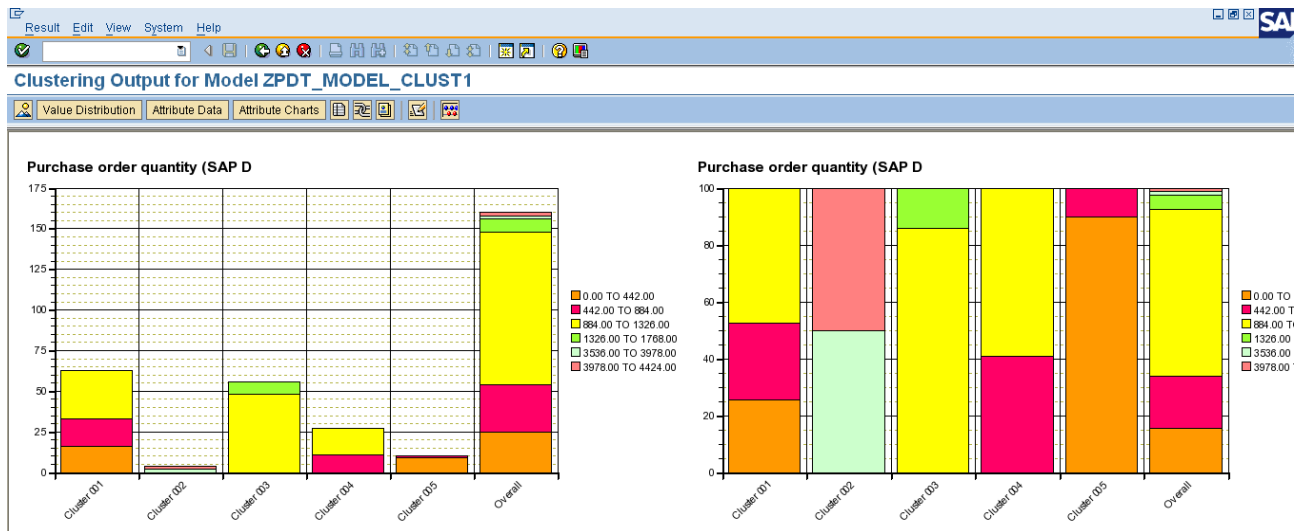
Material

	Cluster 001	Cluster 002	Cluster 003	Cluster 004	Cluster 005	Overall
100-101	1	0	0	0	0	1
100-130	62	4	56	27	10	159

Item number of purchasing docu

	Cluster 001	Cluster 002	Cluster 003	Cluster 004	Cluster 005	Overall
9.00 TO 14.00	11	0	0	0	0	11
19.00 TO 24.00	52	4	0	0	0	56
29.00 TO 34.00	0	0	56	26	10	92
54.00 TO 60.00	0	0	0	1	0	1

- Click on Attribute Charts to “ View Attribute Value Distribution across Clusters”



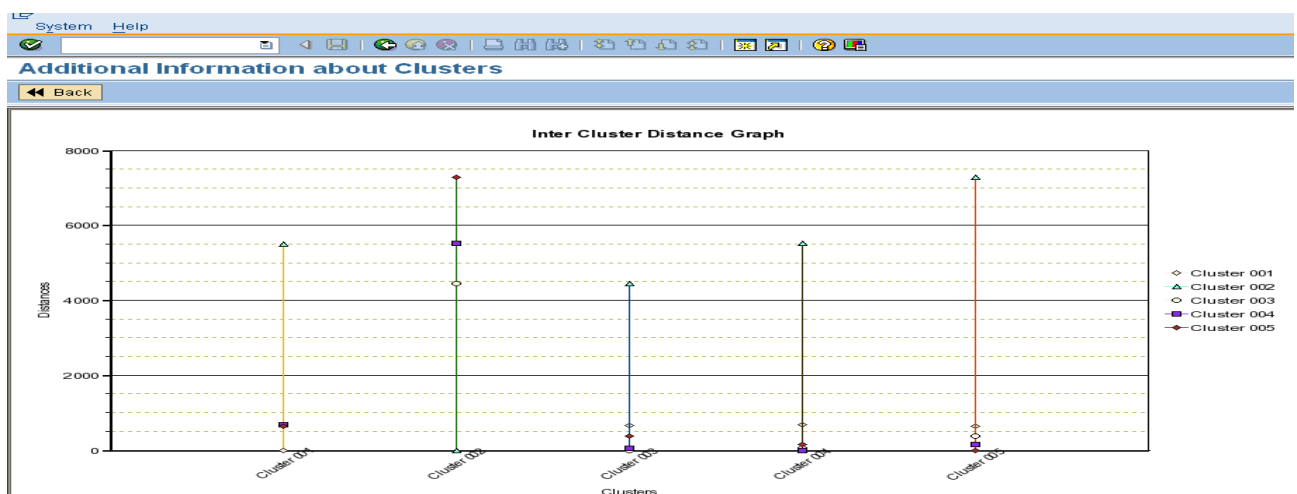
- Click on Intra Cluster Attribute Value Distribution Index to view attribute index

System Help

Attribute Value Index

Cluster ID	Desc	Field Name	Description	Attr.Value (Actual)	Index	Index(Abs)
1	Cluster 001	0D_PO_QTY	Purchase order quantity (SAP D)	0.00 TO 442.00	0,62540	0,62540
1	Cluster 001	0D_PO_QTY	Purchase order quantity (SAP D)	442.00 TO 884.00	0,48878	0,48878
1	Cluster 001	0D_PO_QTY	Purchase order quantity (SAP D)	884.00 TO 1326.00	0,18946-	0,18946
1	Cluster 001	0D_PO_QTY	Purchase order quantity (SAP D)	1326.00 TO 1768.00	1,00000-	1,00000
1	Cluster 001	0D_PO_QTY	Purchase order quantity (SAP D)	3536.00 TO 3978.00	1,00000-	1,00000
1	Cluster 001	0D_PO_QTY	Purchase order quantity (SAP D)	3978.00 TO 4424.00	1,00000-	1,00000

- Click on Inter and Intra Cluster Distance Information



Related Content

[Clustering](#)

[Data Mining](#)

[Data Mining-Decision Trees](#)

For more information, visit the [EDW homepage](#)

Disclaimer and Liability Notice

This document may discuss sample coding or other information that does not include SAP official interfaces and therefore is not supported by SAP. Changes made based on this information are not supported and can be overwritten during an upgrade. SAP will not be held liable for any damages caused by using or misusing the information, code or methods suggested in this document, and anyone using these methods does so at his/her own risk.

SAP offers no guarantees and assumes no responsibility or liability of any type with respect to the content of this technical article or code sample, including any liability resulting from incompatibility between the content within this document and the materials and services offered by SAP. You agree that you will not hold, or seek to hold, SAP responsible or liable with respect to the content of this document.