

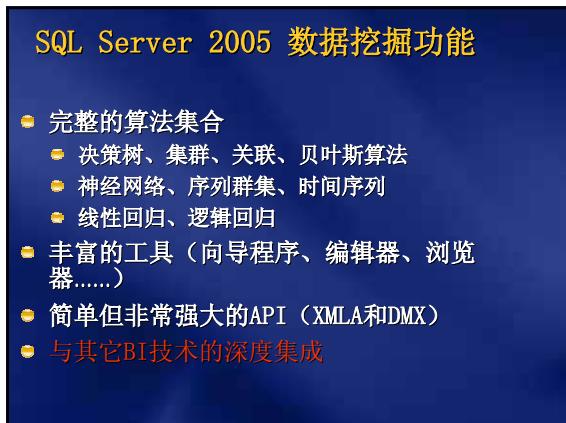


## 议程

- 数据挖掘简介
- Microsoft® SQL Server™ 2005数据挖掘概要介绍
- 利用数据发掘方式使Business Intelligence (BI) 变得更加智能
  - 智能多维数据库
  - 高级ETL
  - 智能报表
- 问题与解答



## 数据挖掘能够完成哪些工作？

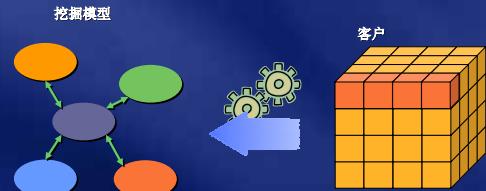


## 智能化多维数据集

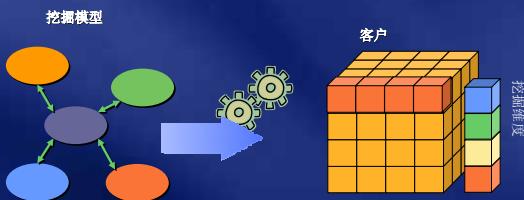
OLAP挖掘应用情境

- 多维数据集 —— 一种语义完善且具有预先计算好的集合值的数据库
  - OLAP数据挖掘应用情景
    - 市场篮分析
    - 客户聚类
    - 客户分类
    - 销售预测
    - ...
  - OLAP挖掘模型
    - 与关系型相同
    - 灵活、快速的建模

OLAP数据挖掘



## OLAP数据挖掘

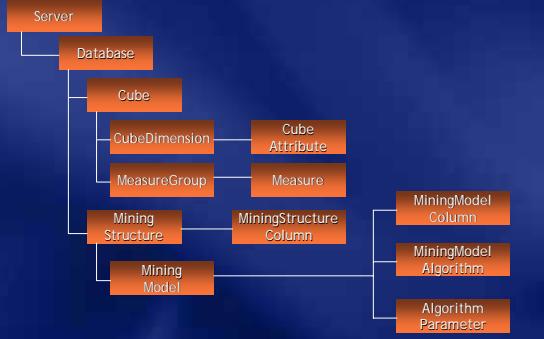


*demo*

挖掘OLAP多维数据集



## AMO对象模型概要介绍



AMO代码

## AMO 代码— 1/4

```
//Connect to database
Server myServer = new Server();
myServer.Connect("DataSource=local host; Catalog=FoodMart");
Database myDatabase = myServer.Databases["FoodMart"];
Cube myCube = myDatabase.Cubes["FoodMart 2000"];
CubeDimension myDimension = myCube.Dimensions["Customer"];

//Create a new mining structure
MiningStructure myMiningStructure =
myDatabase.MiningStructures.Add("CustomerSegment", "CustomerSegment");
myMiningStructure.Source = new CubeDimensionBinding(".", ".",
myCube.ID, myDimension.ID);

//Add gender as an attribute in mining structure
CubeAttribute totalChildren =
myCube.Dimensions["Customer"].Attributes["Total Children"];
ScalarMiningStructureColumn totalChildrenStructureColumn =
Utilities.CreateMiningStructureColumn(totalChildren, false);
myMiningStructure.Columns.Add(totalChildrenStructureColumn);

//Add store sales as an attribute in mining structure
Measure storeSales = myCube.MeasureGroups[0].Measures["Store Sales"];
ScalarMiningStructureColumn storeSalesStructureColumn =
Utilities.CreateMiningStructureColumn(storeSales);
myMiningStructure.Columns.Add(storeSalesStructureColumn);
```

## AMO 代码— 2/4

```
//Connect to database
Server myServer = new Server();
myServer.Connect("DataSource=local host; Catalog=FoodMart");
Database myDatabase = myServer.Databases["FoodMart"];
Cube myCube = myDatabase.Cubes["FoodMart 2000"];
CubeDimension myDimension = myCube.Dimensions["Customer"];
ScalarMiningStructureColumn totalChildrenStructureColumn =
Utilities.CreateMiningStructureColumn(totalChildren, false);
myMiningStructure.Columns.Add(totalChildrenStructureColumn);

//Add store sales as an attribute in mining structure
Measure storeSales = myCube.MeasureGroups[0].Measures["Store Sales"];
ScalarMiningStructureColumn storeSalesStructureColumn =
Utilities.CreateMiningStructureColumn(storeSales);
myMiningStructure.Columns.Add(storeSalesStructureColumn);
```

## AMO 代码— 3/4

```
//Connect to database
Server myServer = new Server();
myServer.Connect("DataSource=local host; Catalog=FoodMart");
Cube myCube = myDatabase.Cubes["FoodMart 2000"];
MiningStructure myMiningModel =
myMiningStructure.CreateMiningModel(true, "CustomerSegment");
myMiningModel.Algorithm =
MiningModelAlgorithm.MICROSOFT_CLUSTERING;

//Create a mining model from the structure
MiningModel myMiningModel =
myMiningStructure.CreateMiningModel(true, "CustomerSegment");
myMiningModel.Algorithm =
MiningModelAlgorithm.MICROSOFT_CLUSTERING;

//Add algorithm parameters
AlgorithmParameter parm1 = new AlgorithmParameter("MODELING_CARDINALITY", 1);
myMiningModel.AlgorithmParameters.Add(parm1);

//Process the model
Try
{
    myMiningStructure.Update(UpdateOptions.ExpandFull);
    myMiningStructure.Process(ProcessType.ProcessFull);
}
catch (Microsoft.AnalysisServices.OperationException e)
{
    ...
}
```

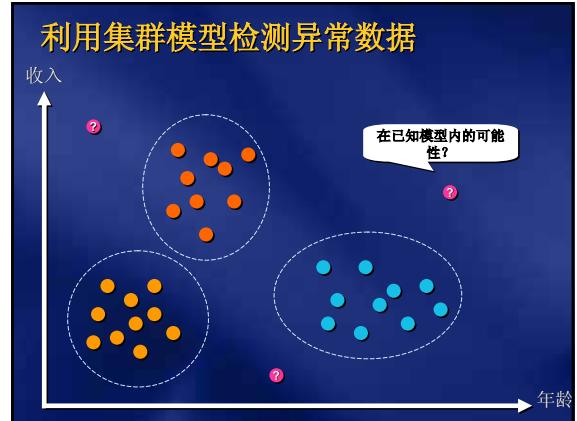
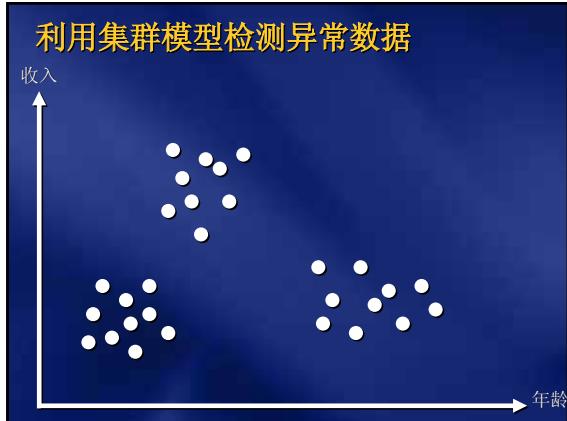
## AMO 代码— 4/4

```
//Create column binding to the cube attributes
public static ScalarMiningStructureColumn CreateMiningStructureColumn(CubeAttribute attribute, bool isKey)
{
    ScalarMiningStructureColumn column = new ScalarMiningStructureColumn();
    column.Name = attribute.Attribute.Name;
    column.IKey = isKey;
    column.Source = new CubeAttributeBinding(attribute.ParentCube.ID,
((CubeDimension)attribute.Parent).ID, attribute.Attribute.ID,
attribute.DataType.Name);
    column.Type =
MiningStructureColumnTypes.GetColumnType(attribute.Attribute.Name,
myColumn.DataType);
    return column;
}
```

## 高级ETL

## 数据挖掘与集成服务

- 数据挖掘任务
  - 处理任务
  - 预测任务
- 数据挖掘转换
  - 预测转换
  - 模型培训转换
  - 文本挖掘
  - 模糊查询
- 数据挖掘准备
  - 清理
  - 采样



PredictCaseLikelihood()

```
Select Customer_Id , PredictCaseLikelihood()
From MySegmentationModel_Prediction Join
OPENQUERY(mydatasource, 'Select Age,
Income
From Customers') As t
On MySegmentationModel_Age = t.Age
And MySegmentationModel_Select Top 10 Customer_Id
From MySegmentationModel_Prediction Join
OPENQUERY(mydatasource, 'Select Age,
Income From Customers') As t
On MySegmentationModel_Age = t.Age
And MySegmentationModel_Income = t.Income
Order by PredictCaseLikelihood() ASC
```

*demo*

用数据挖掘技术进行数据清理  
集成服务与数据挖掘演示

**msdn**

智能报表

服务集成报表生成

- 以DMX作为报表数据源
- 报表设计环境中的预测查询生成器
- 以报表作为数据挖掘预测结果展现方式
- 利用预测分析方式生成由数据驱动的报表
- 预测查询
  - 内容查询
  - 参数化查询
  - 存储过程

