



IDIMT'98:

*Data Mining and
Multidimensional Databases
are the key to Data
Warehouses and WWW.*

*By Michael Haller, Georg Jenichl,
and Josef Küng*

*FAW - Johannes Kepler University Linz
EURECOM - Sophia Antipolis France*



Contents



- ◆ Introduction
- ◆ Data Warehousing
- ◆ Data Mining Technologies
- ◆ Multidimensional Databases
- ◆ A Data Mining Tool (JPCM)
- ◆ A MDDDB-Tool (WebMDDDB)
- ◆ Conclusion



The Price-Data-Circle



Price cut



Huge data

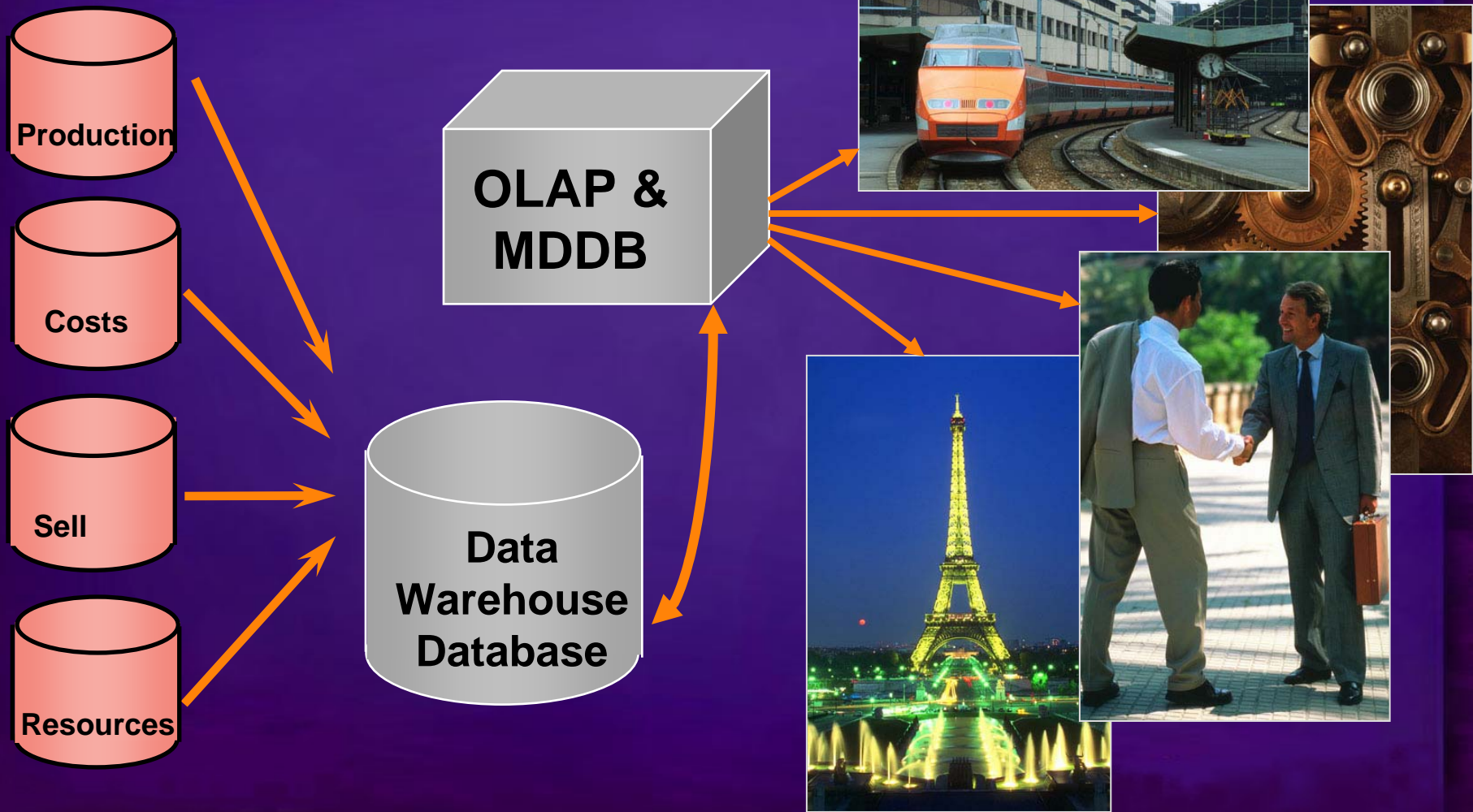


Good Visualization



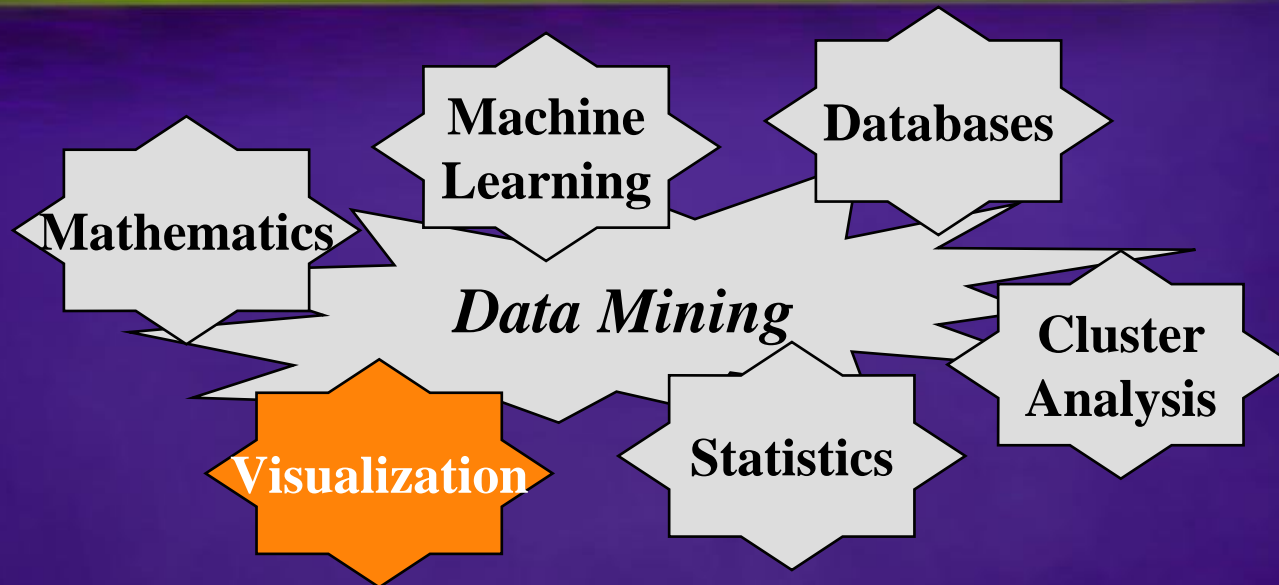
Well done Interpretation

The Process of an Analytical Database





Data Mining ingredients



Visual summaries of data
from a database

Simple, easily interpretable
plots

Linking multiple views of
the same data

Finding unusual data
records



OLTP and OLAP

◆ **OLTP** (= **OnLine Transaction Processing**)

- a lot of users create, modify and cancel records
- optimized for “update”-transaction
- Relational Database

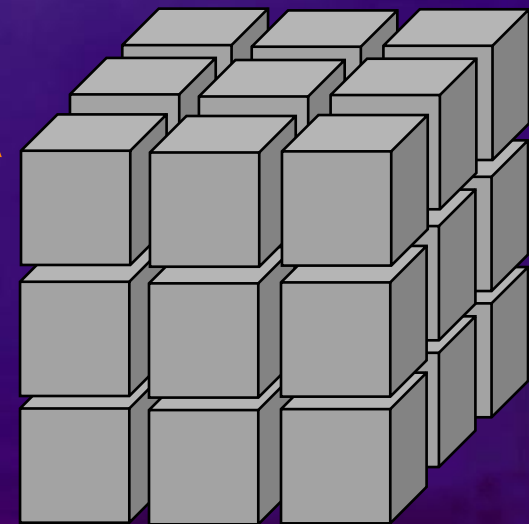
◆ **OLAP** (= **OnLine Analytical Processing**)

- suitable for Ad-Hoc queries
- optimized for analyzes
- Multidimensional Databases

Multidimensional Database



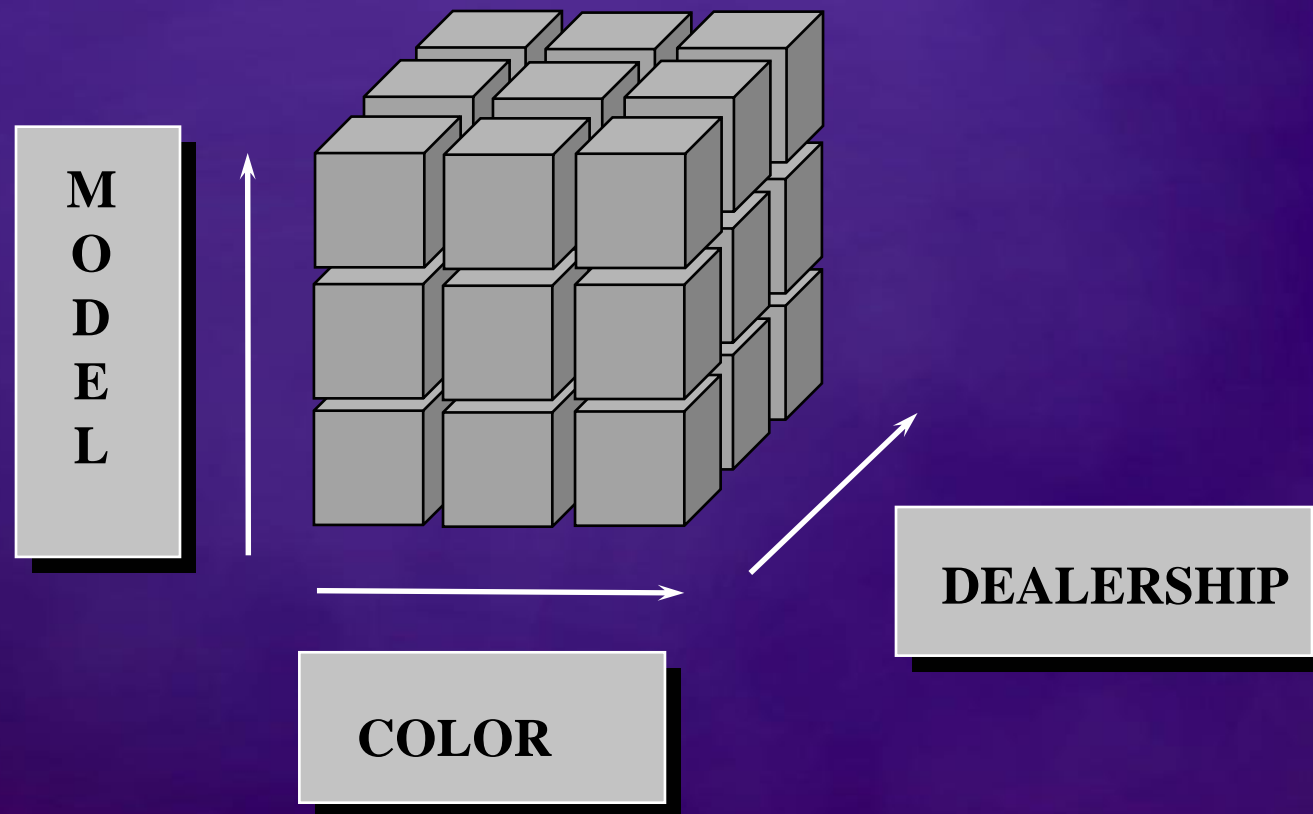
Number	Name	Age
1432	Katharina	25
1421	Hans	27
2211	Sabine	24



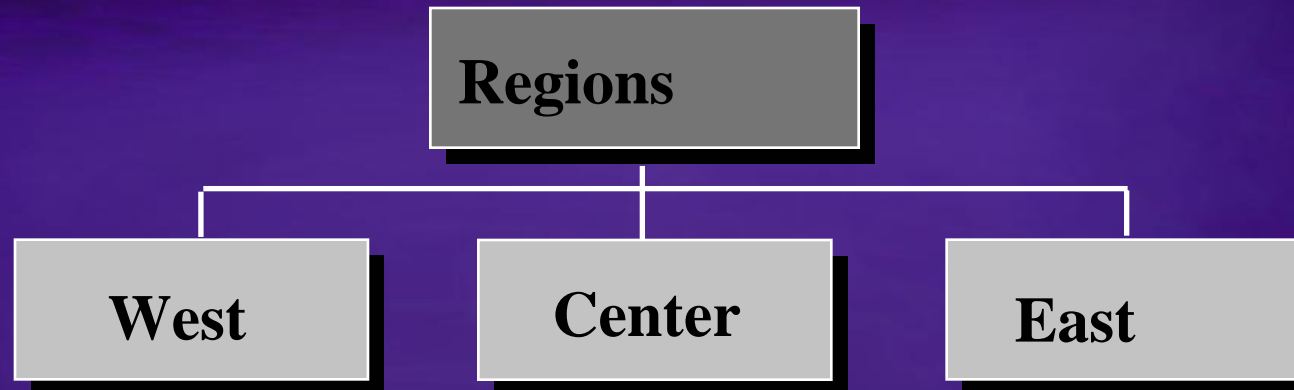
Try to find new techniques
for an efficient data access



Multidimensional Database

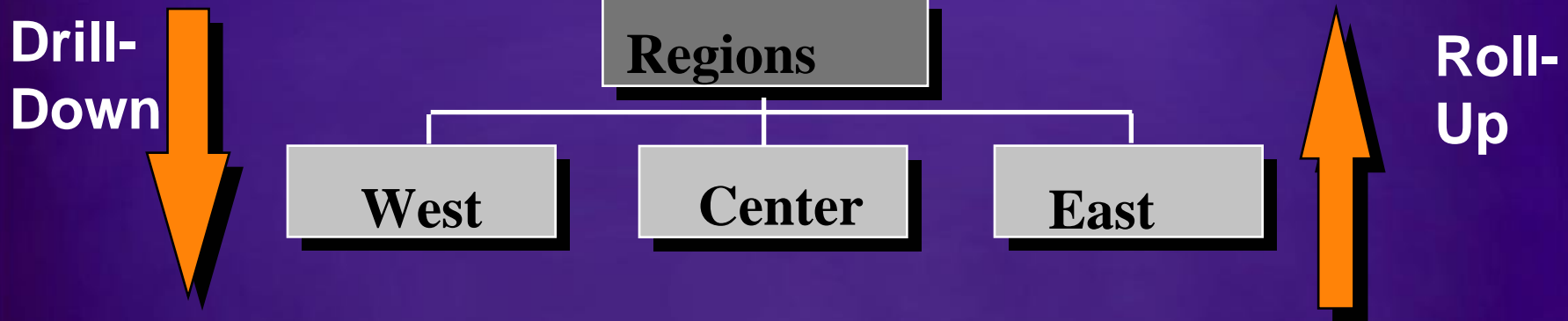


The Hierarchy



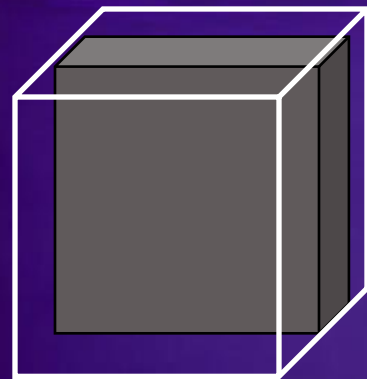
- ◆ Different views for the same data
- ◆ well structured
- ◆ easy way to navigate through the data

Drill-Down and Roll-Up

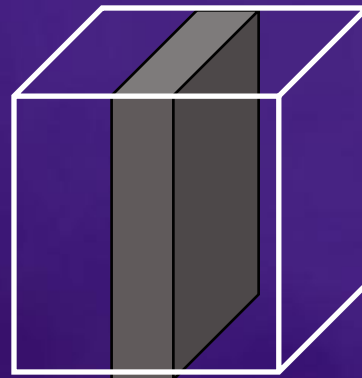


The Navigation Operations are called Drill-Down and Roll-Up

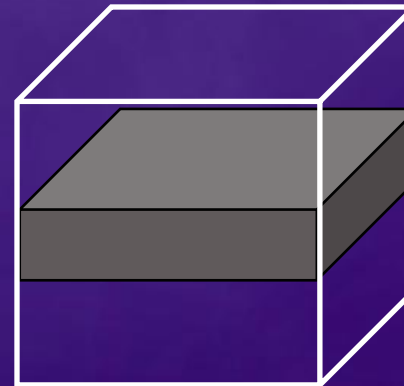
Different Views



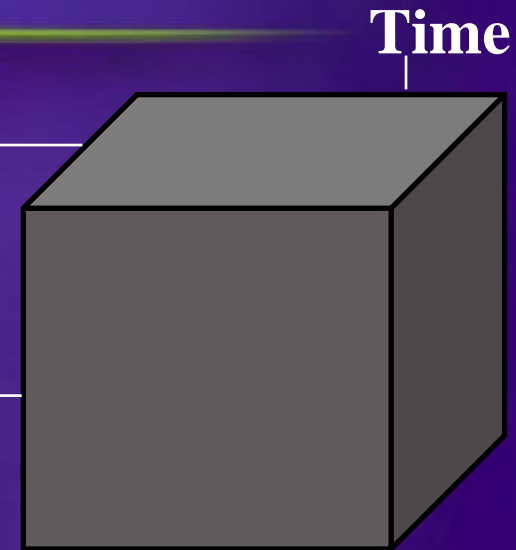
**Product
Manager**



**Treasury
Manager**



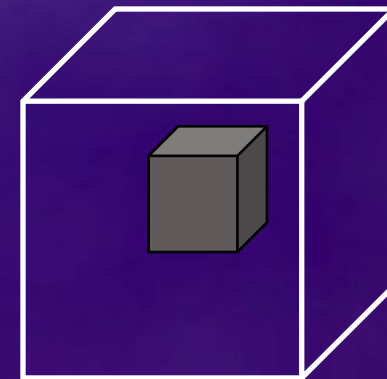
**Regional
Manager**



Product

Market

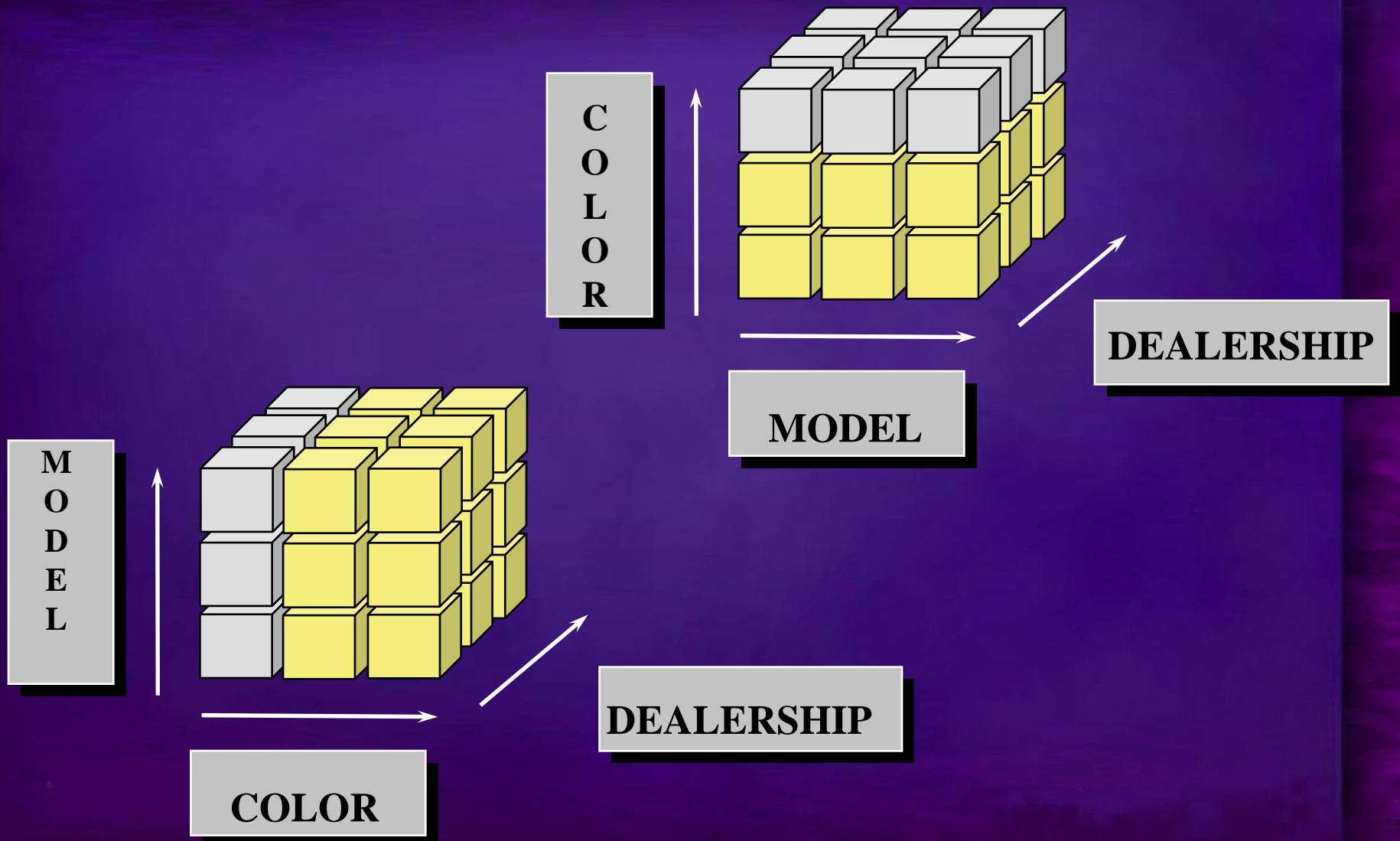
Time



**Ad Hoc
View**



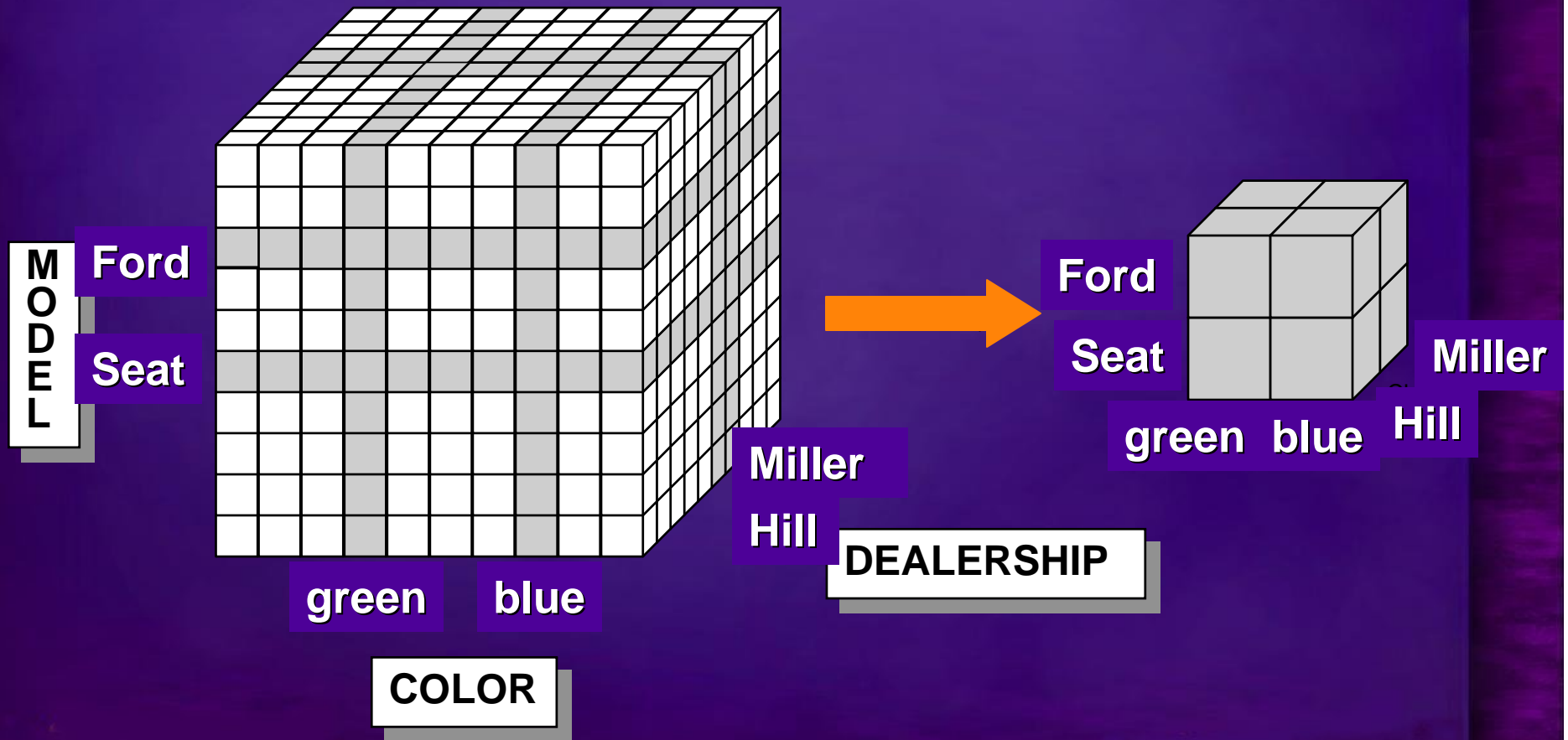
Slice and Dice





Ranging

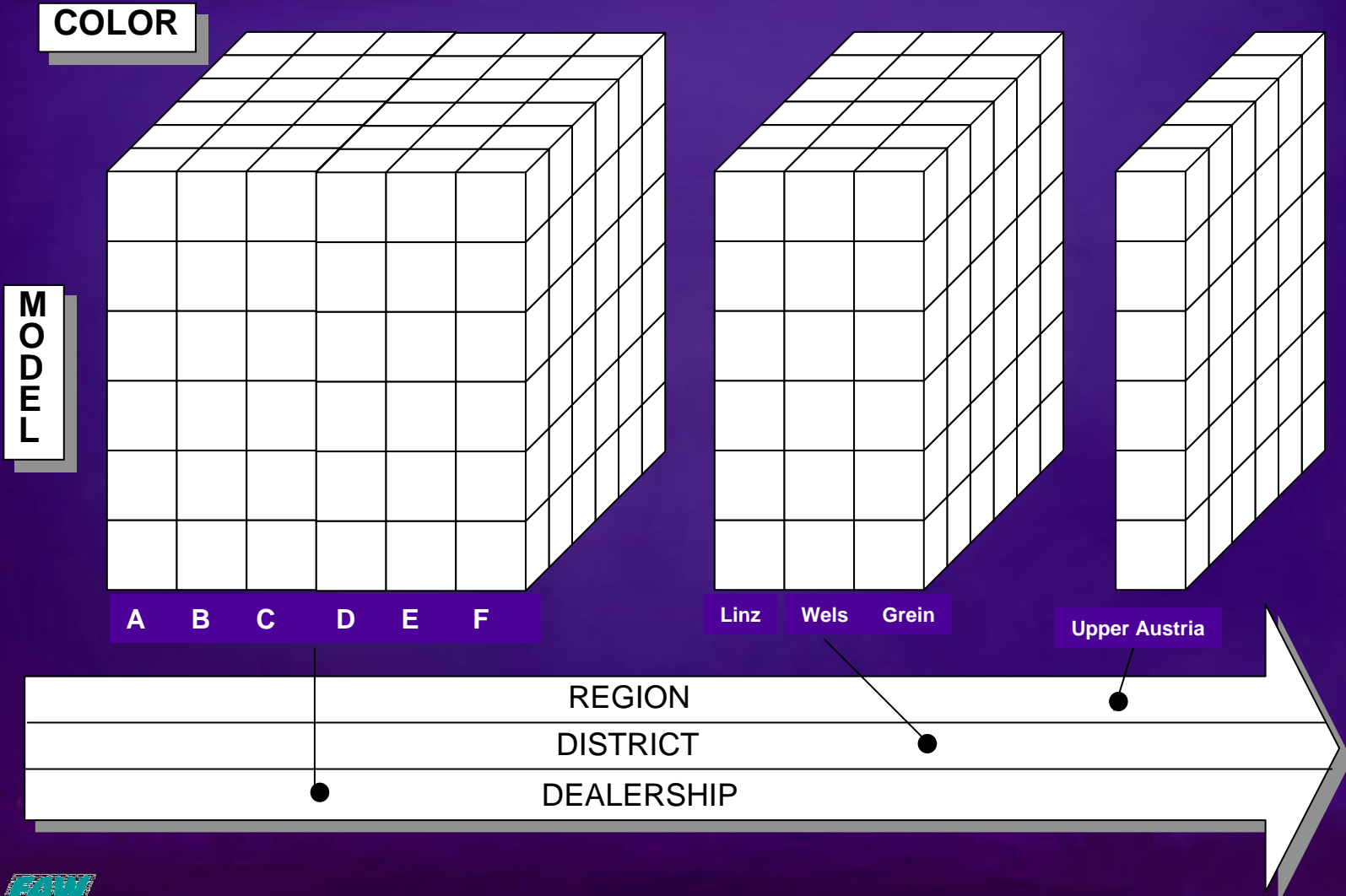
Sales Volumes



The Hierarchy and the Data Cube



Sales Volumes





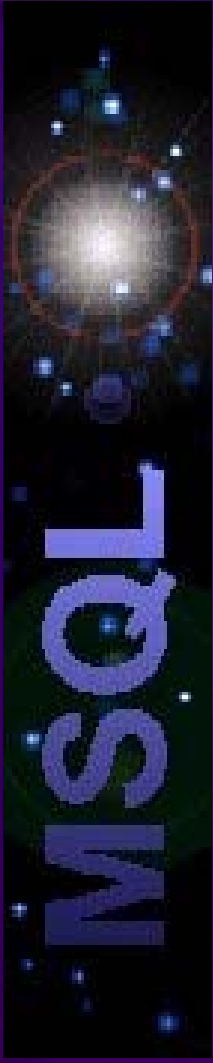
Multidimensional SQL

- ◆ At the moment there is no standard SQL Multidimensional Query Language
- ◆ There are no big differences between the different existing MSOL-languages



MSQL

Multidimensional SQL

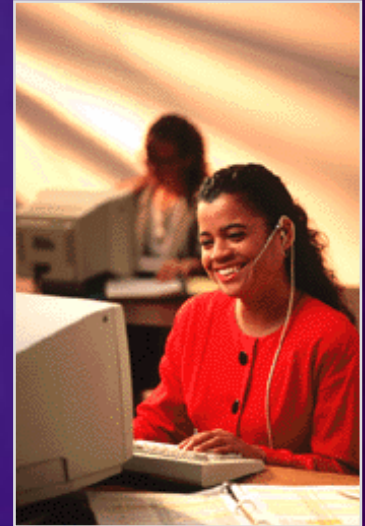


- ◆ Simple Grammar
- ◆ Easy to learn
- ◆ Accessible over the WWW
- ◆ Written in JAVA™

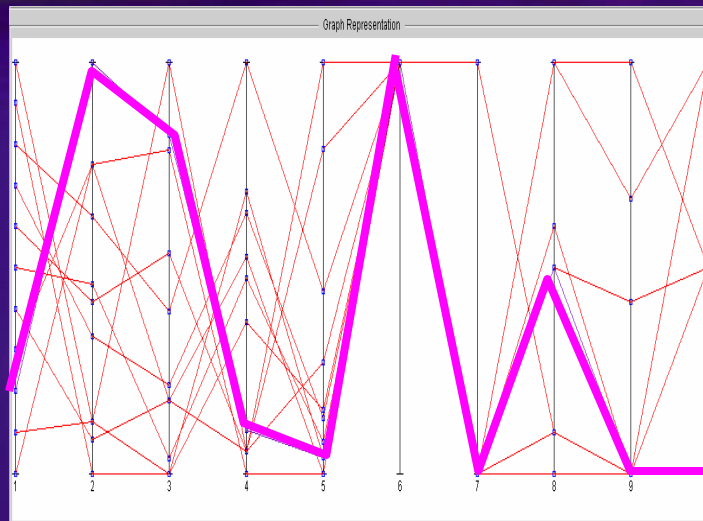
Internet Technology



The Internet



JAVA Parallel Coordinate Miner



JPCM

Visual data mining method

Useful for noticing new phenomena in a dataset

Visualize multidimensional relationships

Detect correlation patterns in multidimensional data



A JPCM - example (1)

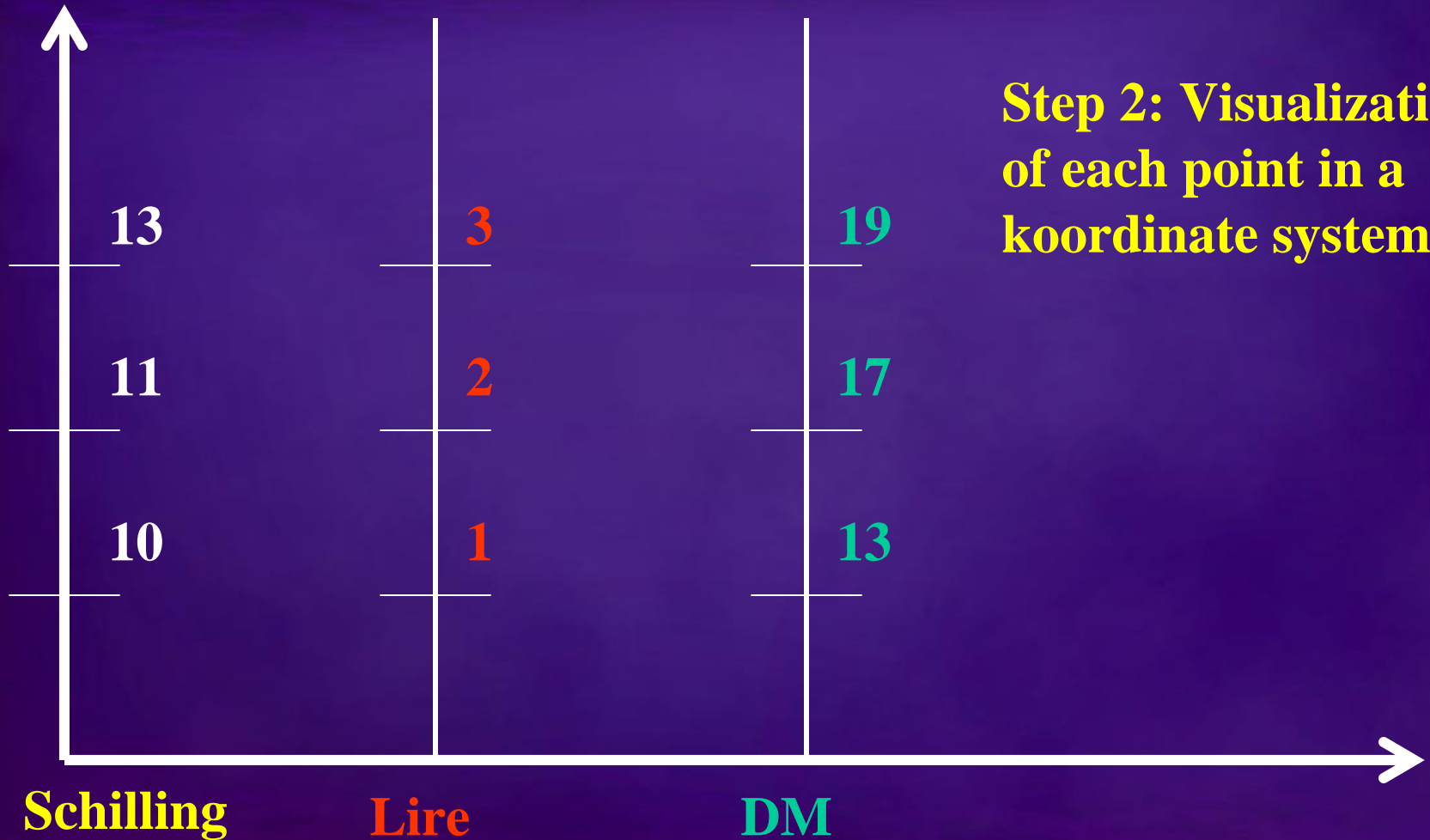
Schilling	Lire	DM	Time
11	3	19	1
10	1	13	2
13	2	17	3

Step 1: Find the minimum and maximum of each attribute

Schilling:	Lire:	DM:	Time:
10	1	13	1
13	3	19	3

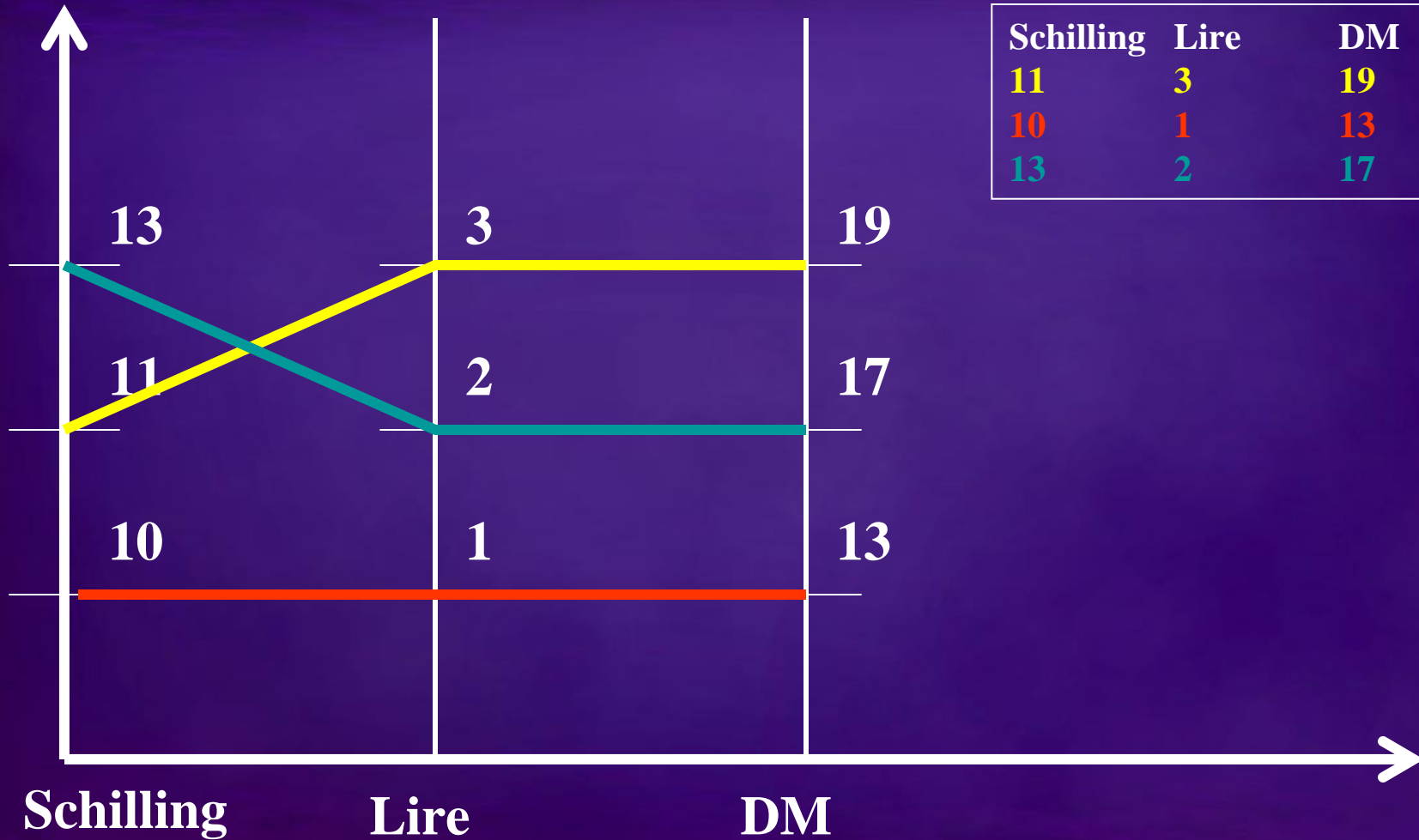


A JPCM - example (2)





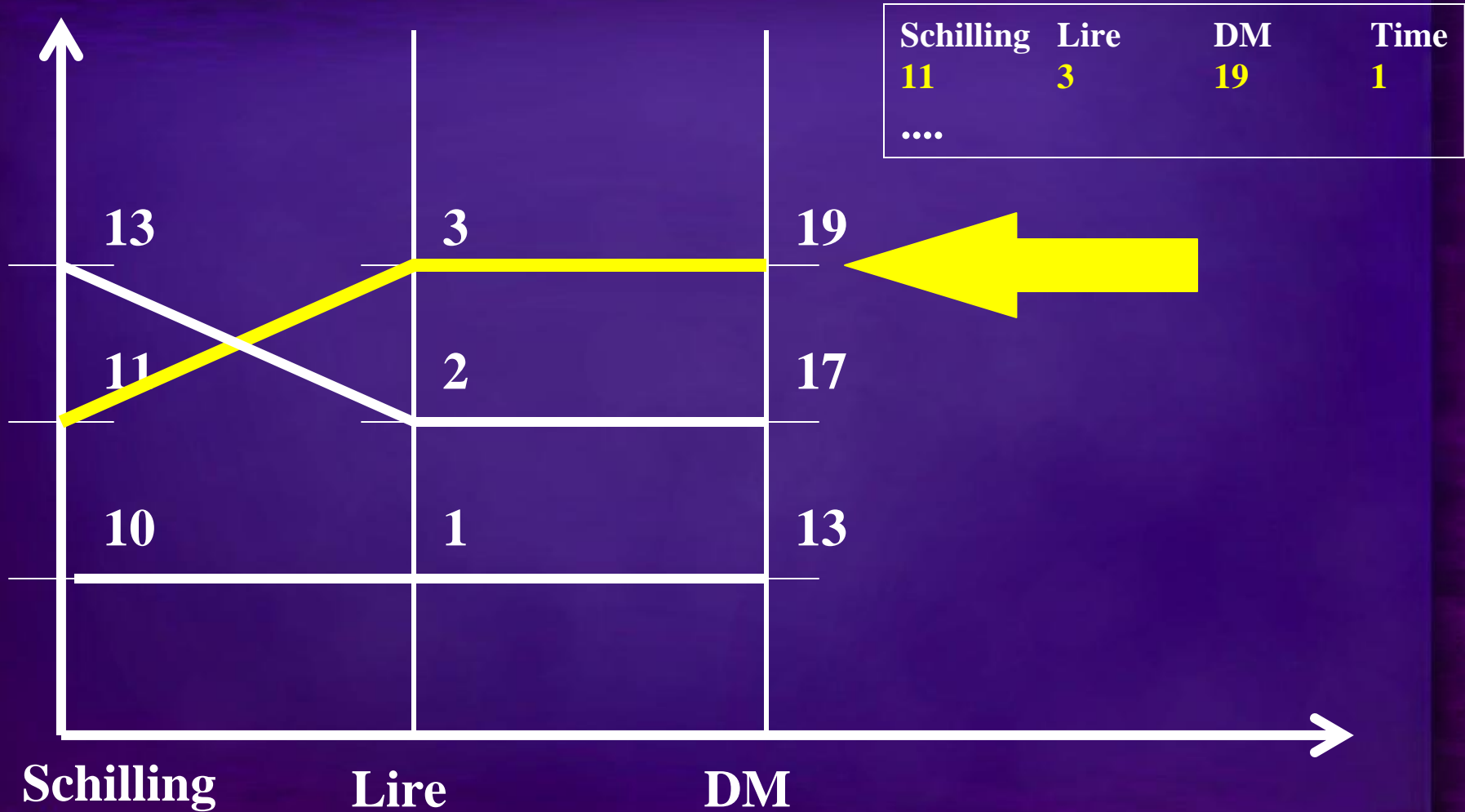
A JPCM - example (3)



Schilling	Lire	DM	Time
11	3	19	1
10	1	13	2
13	2	17	3



A JPCM - example (4)

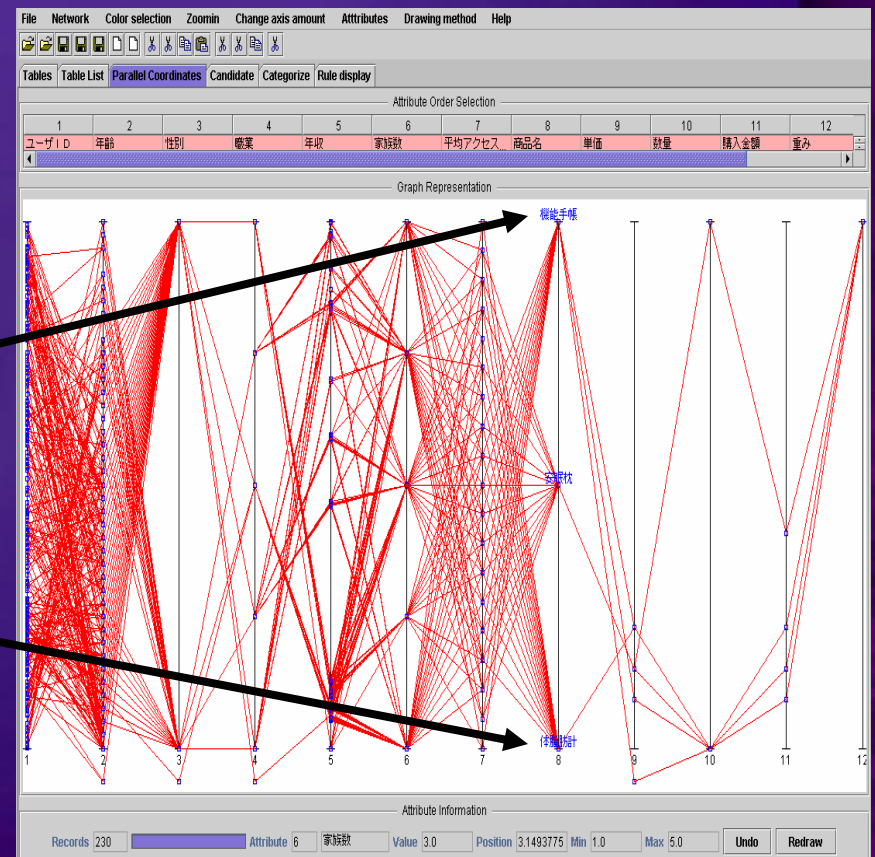
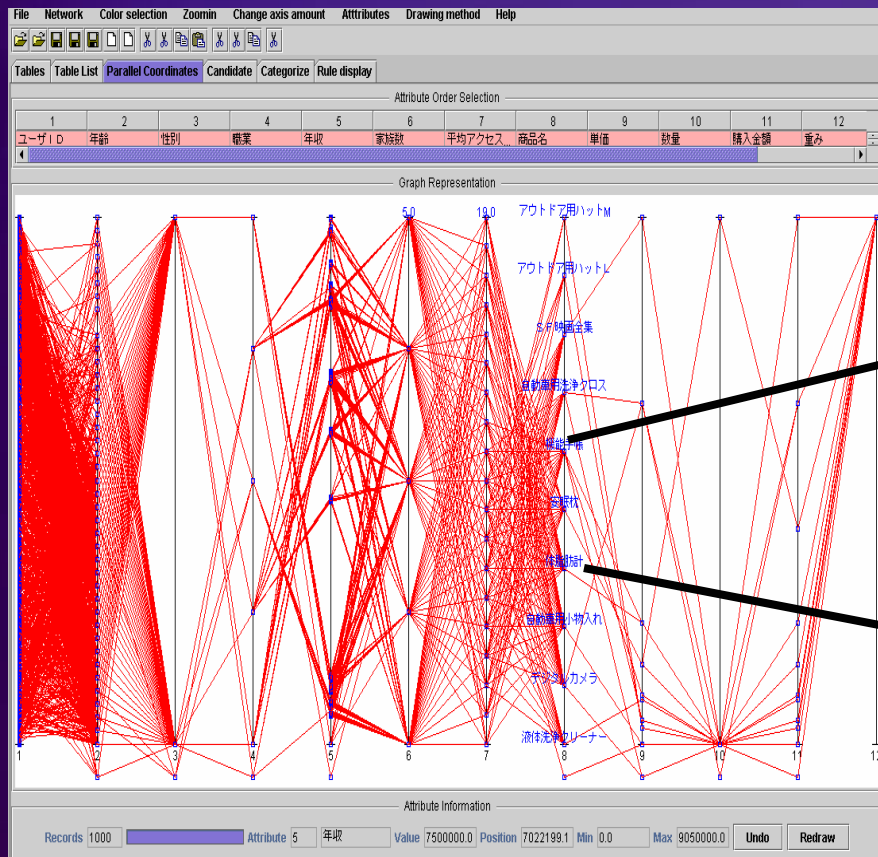




JPCM Table grid representation

ユーザID	年齢	性別	職業	年収	家族数	平均アクセ...	商品名	単価	数量	購入金額	重み
2933.0	34.0	男	公務員	7870000.0	2.0	12.0	液体洗浄ク...	11845.0	1.0	11845.0	1.0
6671.0	NULL	男	会社員	7580000.0	4.0	10.0	デジタルカ...	46659.0	1.0	46659.0	1.0
4704.0	NULL	男	会社員	6270000.0	4.0	11.0	デジタルカ...	46659.0	1.0	46659.0	1.0
7077.0	19.0	男	学生	550000.0	1.0	15.0	自動車用小...	6695.0	1.0	6695.0	1.0
1540.0	38.0	男	会社員	6290000.0	1.0	8.0	自動車用小...	6695.0	1.0	6695.0	1.0
7742.0	18.0	男	学生	520000.0	4.0	16.0	自動車用小...	6695.0	1.0	6695.0	1.0
7490.0	19.0	男	学生	540000.0	1.0	12.0	自動車用小...	6695.0	1.0	NULL	1.0
5075.0	41.0	男	自営業	8810000.0	3.0	7.0	デジタルカ...	46659.0	1.0	46659.0	1.0
4564.0	19.0	男	学生	480000.0	4.0	9.0	自動車用小...	6695.0	1.0	6695.0	1.0
6167.0	24.0	男	会社員	5390000.0	4.0	13.0	デジタルカ...	46659.0	NULL	46659.0	1.0
2842.0	39.0	女	主婦	0.0	4.0	11.0	体脂肪計	20909.0	1.0	20909.0	1.0
5061.0	31.0	男	会社員	6290000.0	3.0	7.0	デジタルカ...	NULL	1.0	46659.0	1.0
4851.0	37.0	男	自営業	7630000.0	5.0	13.0	デジタルカ...	46659.0	1.0	46659.0	1.0
6762.0	22.0	男	学生	730000.0	1.0	12.0	機能手帳	20909.0	1.0	20909.0	1.0
6202.0	18.0	男	学生	530000.0	3.0	13.0	自動車用小...	6695.0	1.0	6695.0	1.0
2884.0	32.0	女	主婦	0.0	3.0	11.0	体脂肪計	20909.0	1.0	20909.0	1.0
3906.0	21.0	男	学生	680000.0	1.0	8.0	デジタルカ...	46659.0	1.0	46659.0	1.0
6118.0	23.0	女	会社員	4240000.0	4.0	12.0	体脂肪計	20909.0	1.0	20909.0	1.0
6447.0	37.0	男	公務員	7840000.0	2.0	16.0	デジタルカ...	46659.0	1.0	46659.0	1.0
2037.0	19.0	女	学生	490000.0	3.0	7.0	安眠枕	15965.0	1.0	15965.0	1.0
2310.0	19.0	男	学生	550000.0	1.0	3.0	自動車用小...	6695.0	1.0	6695.0	1.0
4760.0	18.0	男	学生	570000.0	1.0	12.0	自動車用小...	6695.0	1.0	6695.0	1.0
2338.0	26.0	男	会社員	5350000.0	5.0	3.0	機能手帳	12360.0	1.0	12360.0	1.0
4508.0	22.0	男	学生	710000.0	4.0	9.0	デジタルカ...	46659.0	1.0	46659.0	1.0
4207.0	38.0	女	主婦	0.0	3.0	12.0	体脂肪計	20909.0	1.0	20909.0	1.0
6475.0	52.0	男	公務員	8820000.0	3.0	17.0	安眠枕	15965.0	2.0	31930.0	1.0
4718.0	32.0	男	会社員	6220000.0	2.0	11.0	デジタルカ...	46659.0	1.0	46659.0	1.0

JPCM range zoom-in



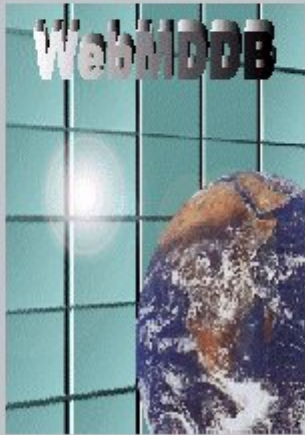


WebMDDB

Applet Viewer: WebMDDB.class

Applet

Connect | Tables | Columns | Dimensions | View | Chart3D | Chart2D | MSQL



Connection Parameters

Server:

Username:

Password:

Connect String:

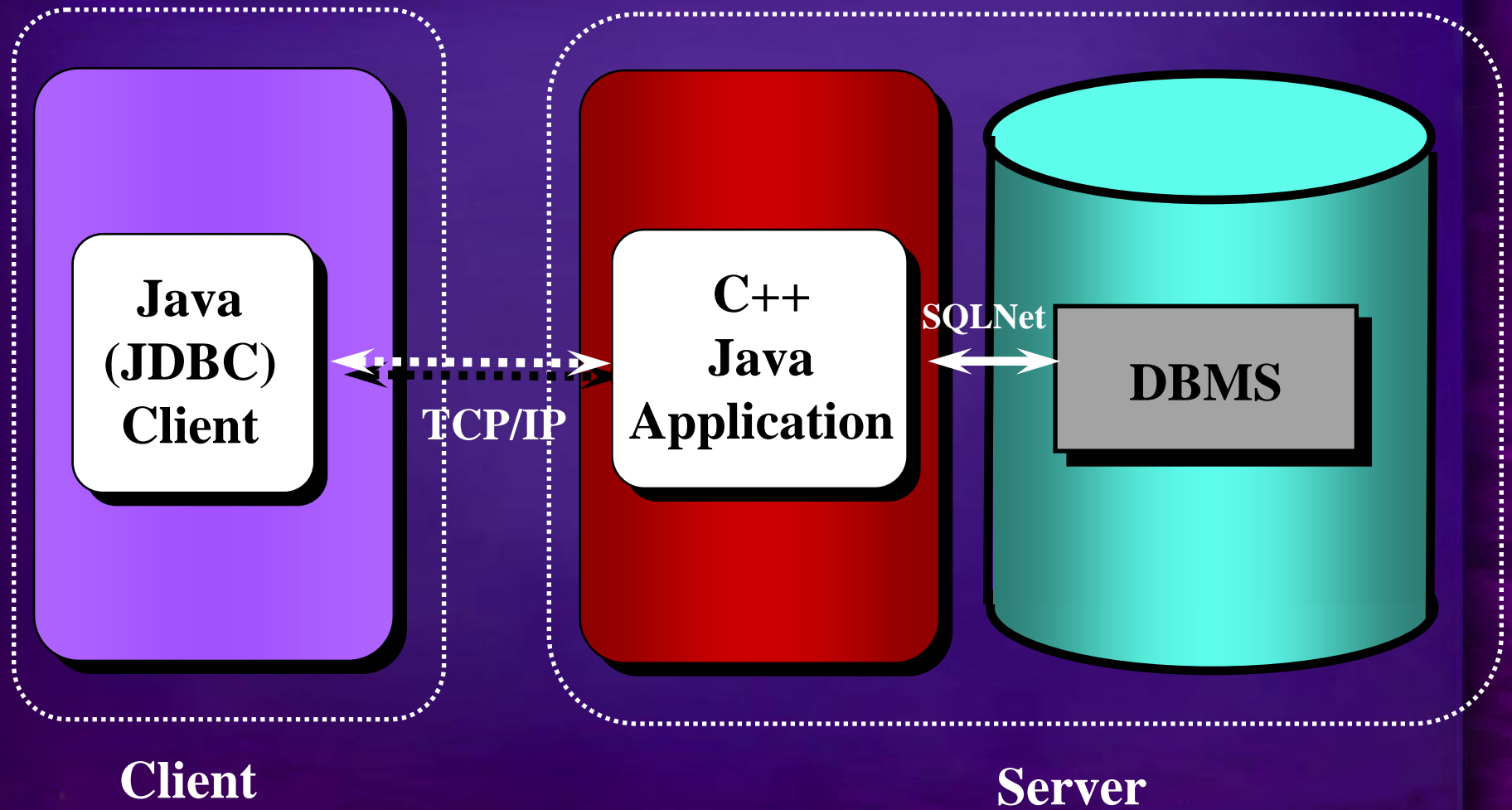
Connect Disconnect Help

WebMDDB - Proto (October, 1997)

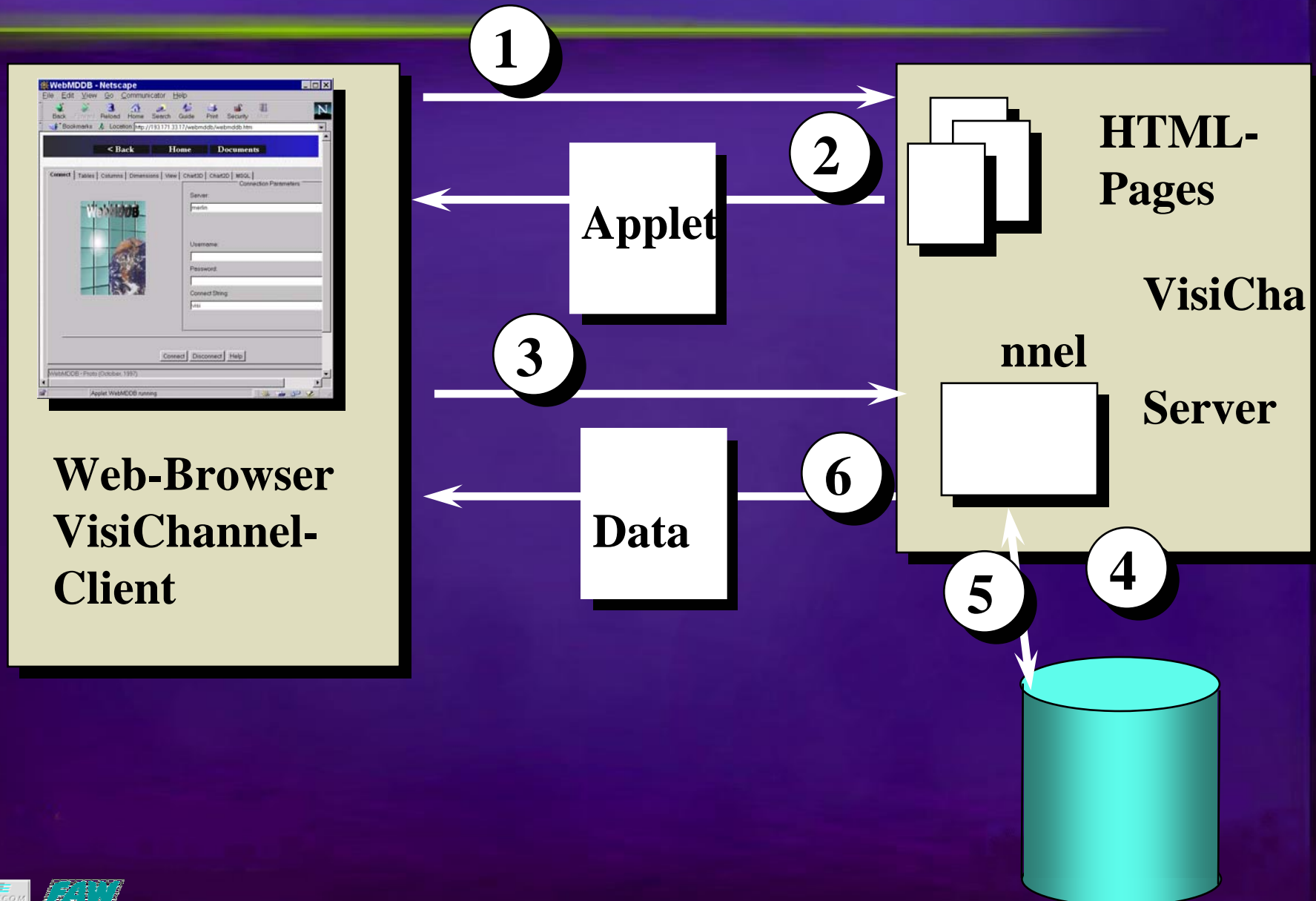
Applet started.



The 3-tier Architecture

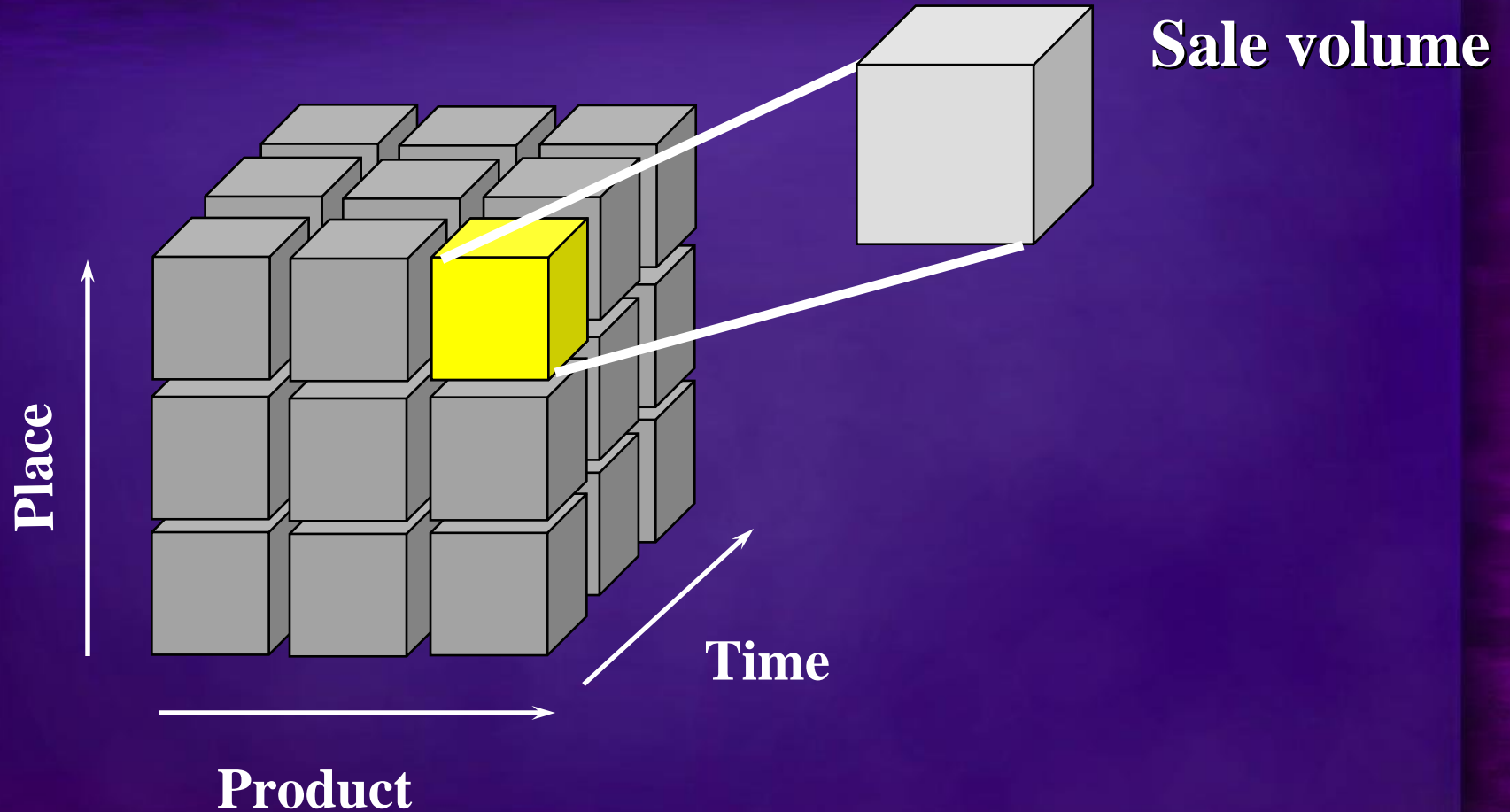


How does WebMDDDB work?





An example - The data cube



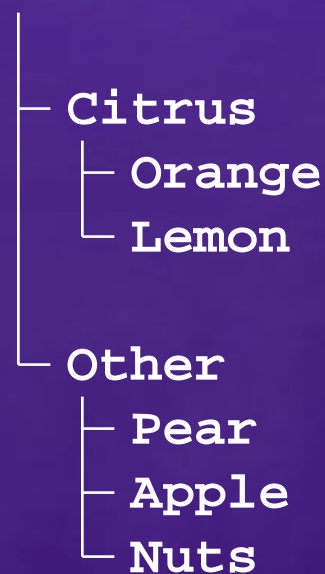
The hierarchy in the example



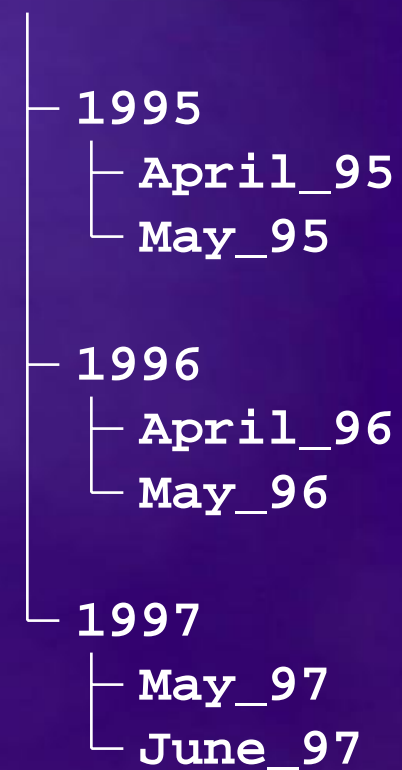
Region



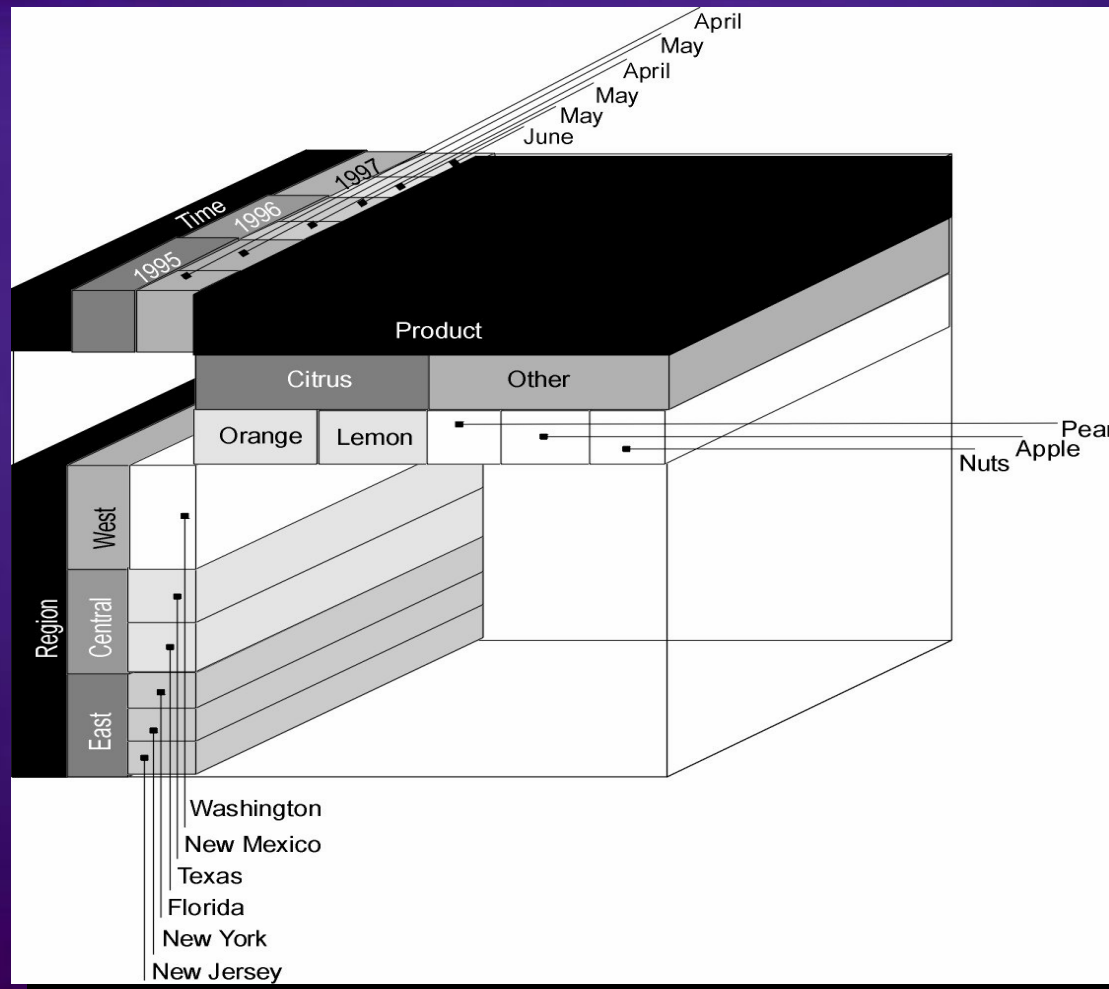
Product



Time



The 3D - Data Cube



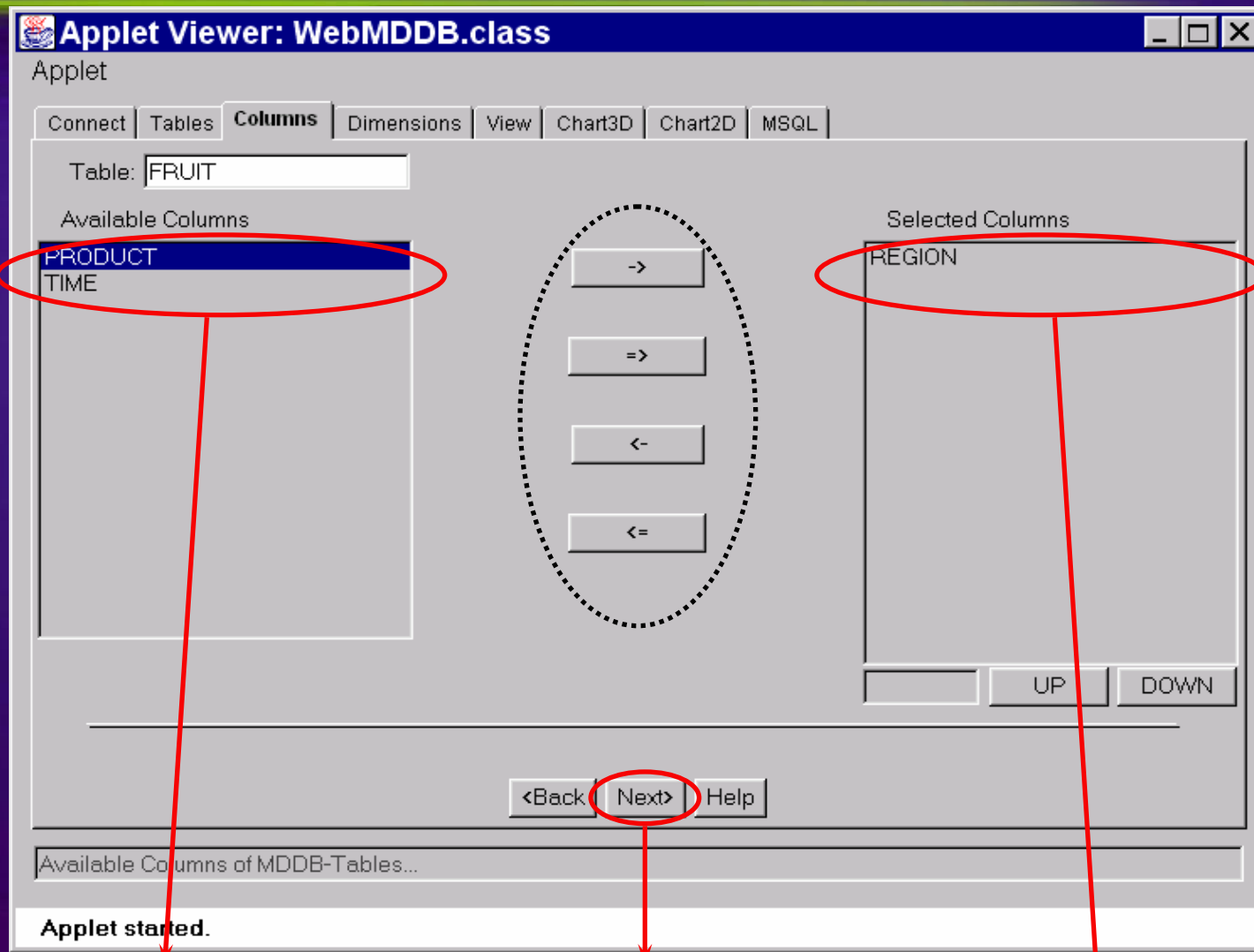
Select a data cube



The screenshot shows the 'Applet Viewer: WebMDDB.class' window. The 'Tables' tab is active, displaying a tree view of 'Server-Tables' with two sub-items: 'FRUIT' and 'EMP'. The 'FRUIT' item is highlighted with a blue background. Below the tree view, the 'Selected Table' field contains the text 'FRUIT'. At the bottom of the window, there are buttons for '<Back', 'Next>', and 'Help'. The status bar at the very bottom indicates 'Applet started.'

The list of all available data cubes.

Selection of the dimensions



Available dimensions

Next...

Selected dimensions



The 2D - Visualization

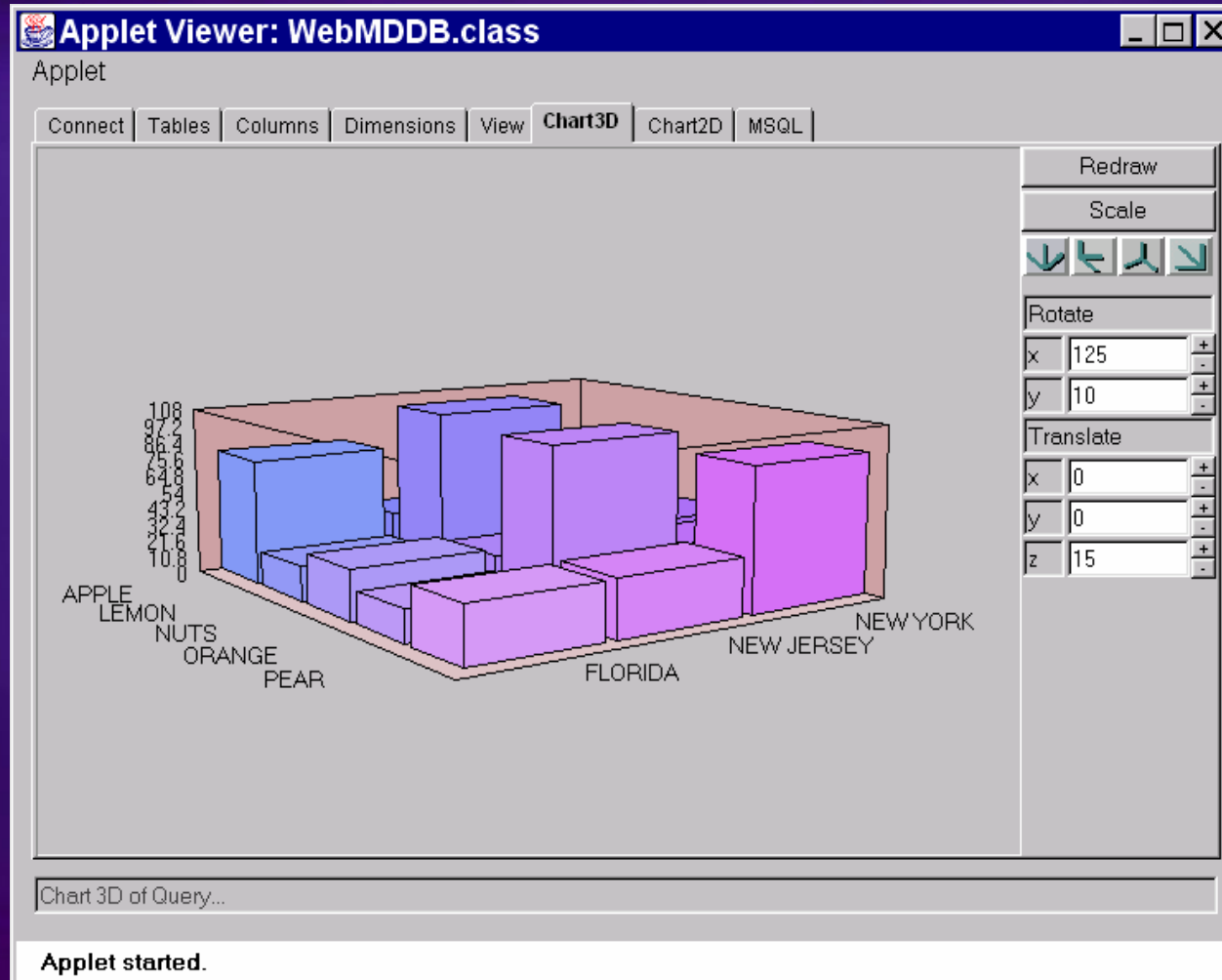
The panel with all available dimensions for the navigate-operation

The screenshot shows the 'Applet Viewer: WebMDDB.class' window. The 'View' tab is selected, displaying a table titled 'The 2D View'. On the left, a 'Dim' panel lists available dimensions: REGION, PRODUCT, TIME, and several 'none' options. The table displays data for various regions across four product categories: APPLE, LEMON, NUTS, and ORANGE.

	APPLE	LEMON	NUTS	ORANGE
FLORIDA	128	75	82	72
NEW JERSEY	76	225	79	289
NEW MEXICO	141	271	198	209
NEW YORK	35	172	97	143
TEXAS	66	164	71	78
WASHINGTON	67	81	86	79

A "Grid View" - Table

The 3D - Visualization





Drill-Down & Roll-Up

Applet Viewer: WebMddb.class

Applet

Connect | Tables | Columns | **Dimensions** | View | Chart3D | Chart2D | MSQL

Dimension for drill down & roll up...

```
graph TD; DH[Dimension-Hierarchy] --> REGION[REGION]; REGION --> EAST[EAST]; REGION --> CENTRAL[CENTRAL]; CENTRAL --> TEXAS[TEXAS]; CENTRAL --> NM[NEW MEXICO];
```

Selected Dimension: CENTRAL

<Back | Generate | Help

Change the Dimension...

Applet started.

Slice and Dice



Applet Viewer: WebMddb.class

Applet

Connect Tables Columns Dimensions **View** Chart3D Chart2D MSQL

Dim

REGION

PRODUCT

TIME

none

none

none

none

none

none

none

none

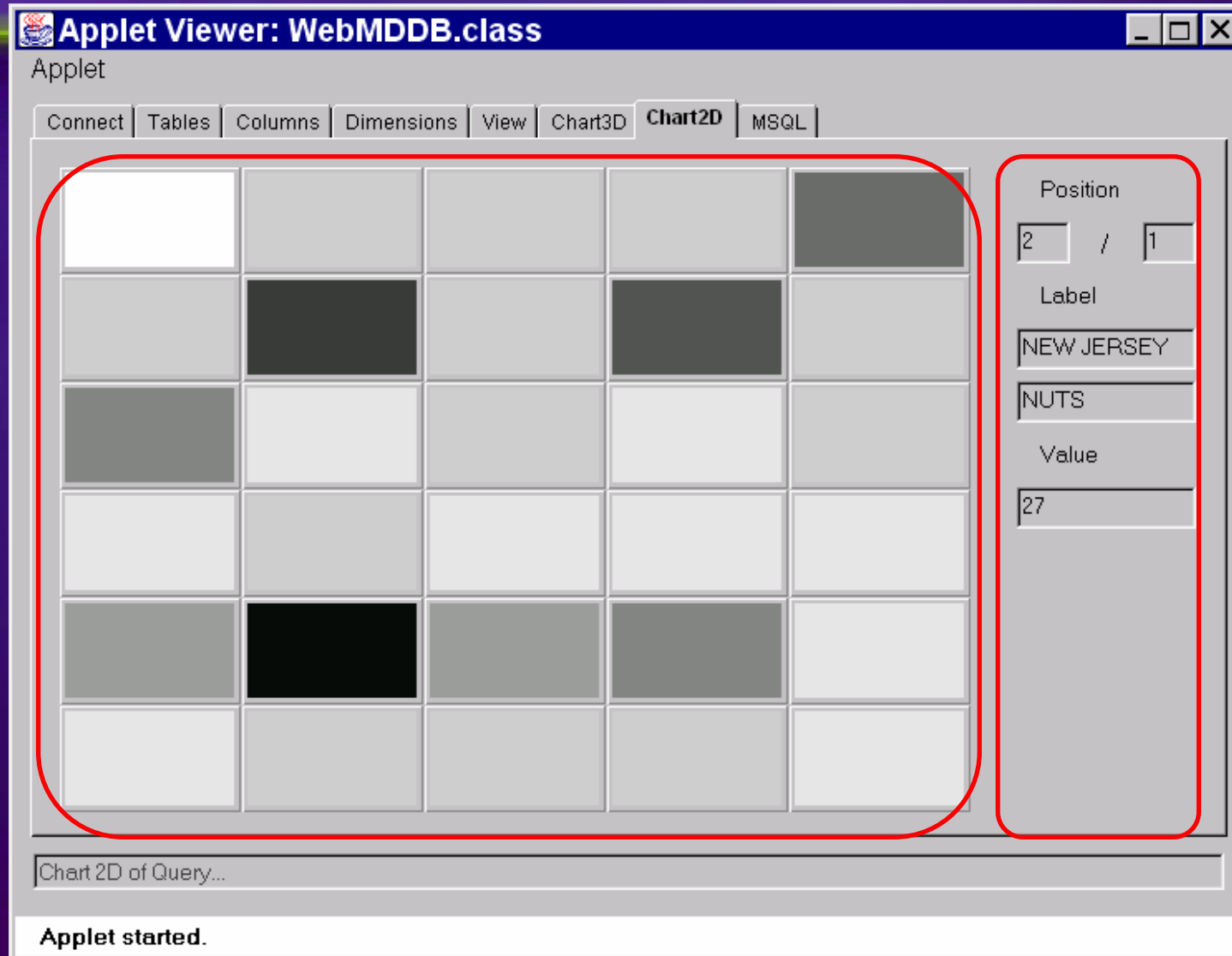
The 2D View

	APPLE	LEMON	NUTS	ORANGE
NEW YORK	35	172	97	143
NEW JERSEY	76	225	79	289
FLORIDA	128	75	82	72

View of Query...

Applet started.

The distribution of the data



The MSQL-Panel



Applet Viewer: WebMDDB.class

Applet

Connect Tables Columns Dimensions View Chart3D Chart2D **MSQL**

MSQL statement..

```
select sal from fruit, product p, region r, time t
where p.product == 'PRODUCT',
      r.region == 'REGION',
      t.time == 'TIME';
```

Parsing...
Calculating...

	ORANGE	LEMON	NUTS	APPLE	PEAR
NEW YORK	143	172	97	35	158
NEW JERSEY	289	225	79	76	126

Execute Clear Input Clear Output Help

Insert a MSQL Query...

Applet started.

MSQL - Example 1



Get all product-sales of all regions from the whole period of time.

```
SELECT sal FROM fruit,  
        product p, region r, time t  
  
WHERE p.product == 'PRODUCT',  
       r.region == 'REGION',  
       t.time == 'TIME';
```

MSQL - Example 2



Get all product-sales of all eastern region from the last year (1997).

```
SELECT sal FROM fruit,  
        product p, region r, time t  
  
WHERE p.product == 'PRODUCT',  
      r.region == 'EAST',  
      t.time == 'T1997';
```

MSQL - Example 3



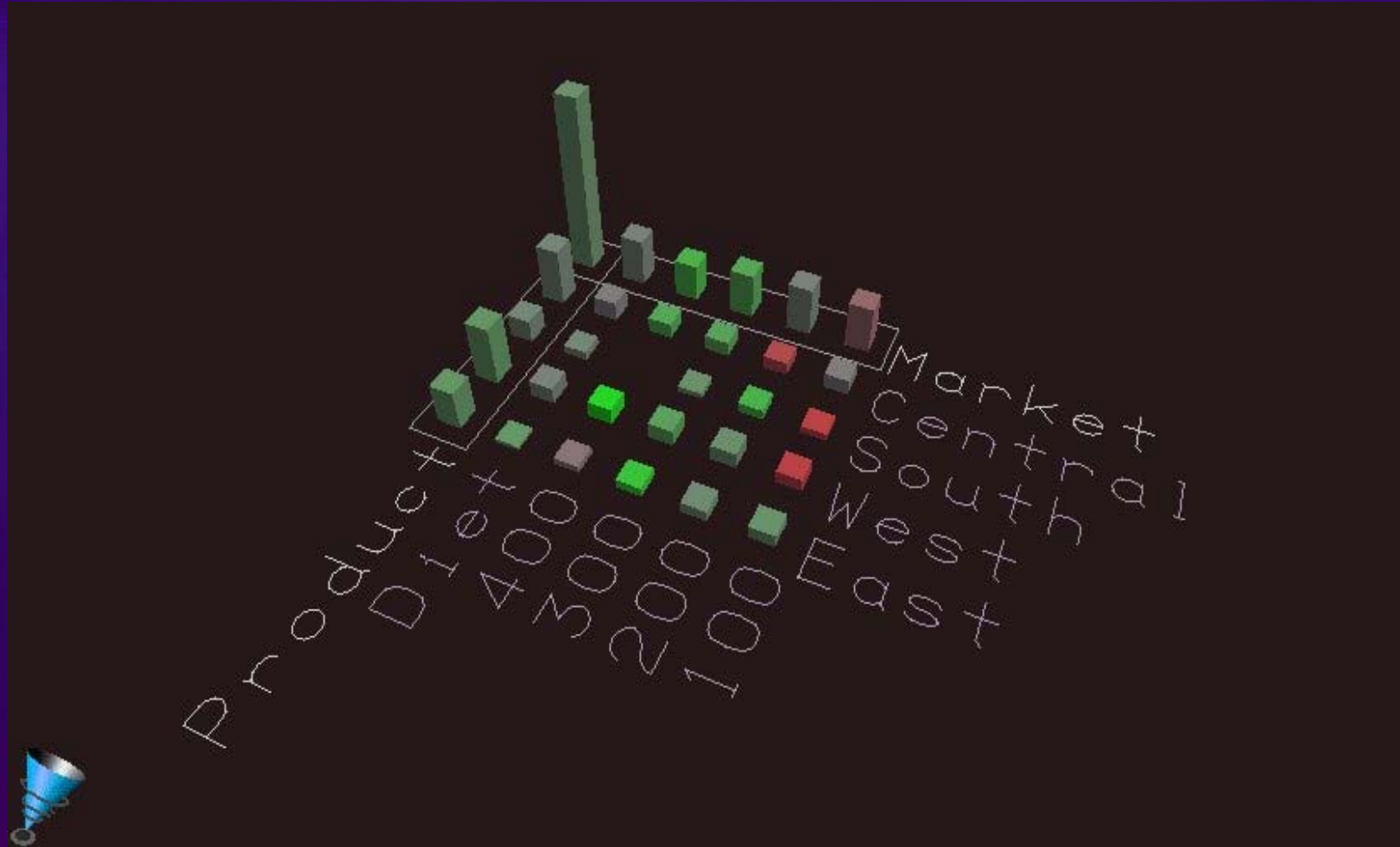
Get the sale for all citrus-fruits excepted the orange and all non-citrus-fruits excepted the pears from Texas and all eastern-regions of the last year (1997).

```
SELECT sal FROM fruit,  
        product p, region r, time t
```

```
WHERE p.product == 'CITRUS' AND p.product != 'ORANGE',  
      p.other == 'OTHER' AND p.other != 'PEAR',  
      r.region == 'EAST' AND r.central == 'TEXAS',  
      t.time == 'T1997'
```



Conclusion - VRML





Questions?

