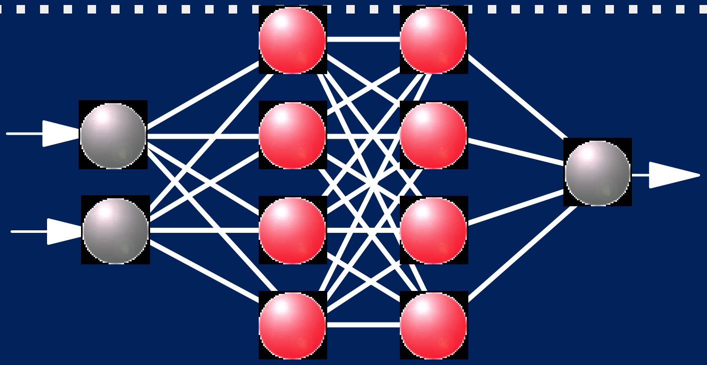
DB2 Data Management Software

IBM

e business software



DB2 Intelligent Miner for Data and DB2 Intelligent Miner Scoring, Modeling, & Visualization

Gregor Meyer
gregorm@us.ibm.com

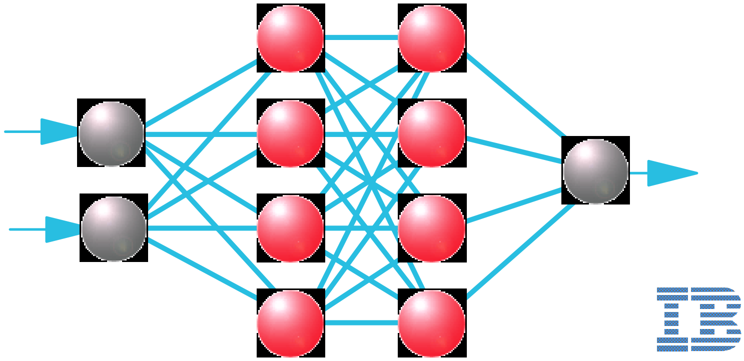
IBM Software Group

IBM Confidential

Note: the conference call won't cover all of the following pages in detail, some are just for your reference.

Agenda

- Data Mining, quick overview
  - Concepts, Business uses
- DB2 Intelligent Miner Technology
  - Intelligent Miner for Data, the workbench
  - IM Scoring, Modeling, Visualization
- Trends, directions
  - Are we on the right track?

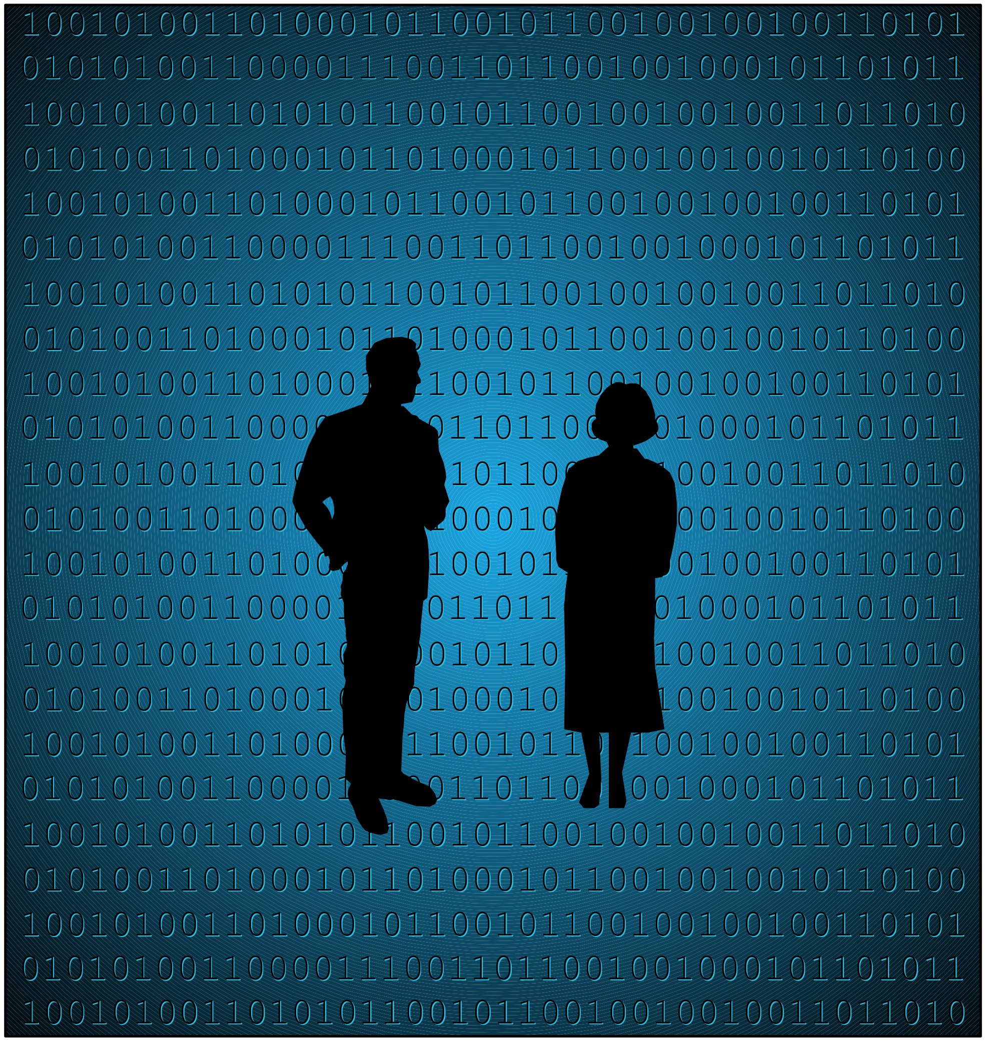


We've got plenty of data. What we need are answers.

find important attributes

personalization

Marketing automation



1-to-1 Marketing

give the right message to the right person

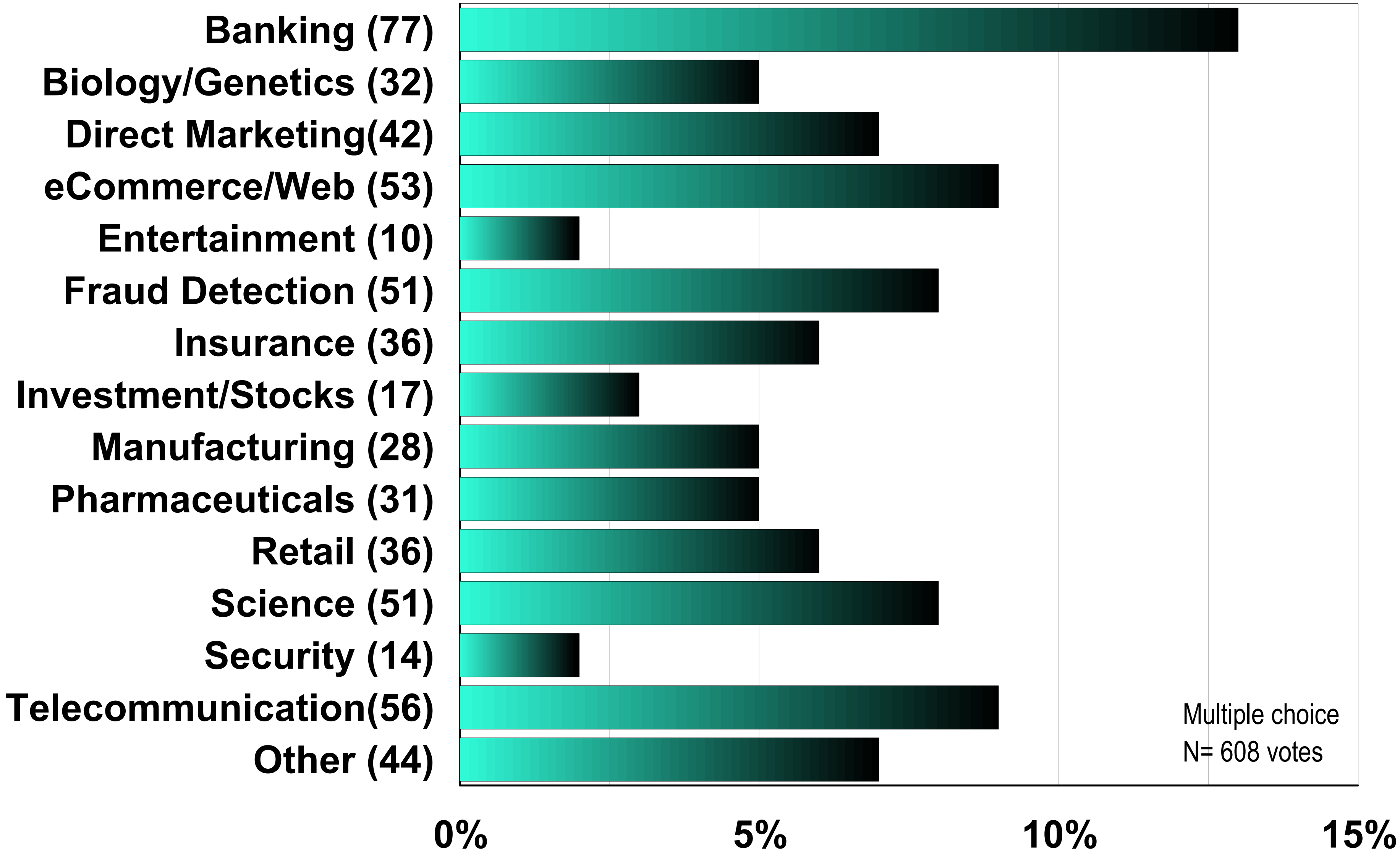
what is our risk?

predict customer behavior

detect anomalies

quality analysis

Where do you plan to use data mining in 2002?



source: KDnuggets : Polls : Data Mining Applications in June 2002
http://www.kdnuggets.com/polls/current_application_fields_2002.htm

DB2 Data Management Software

IBM

Primary Data Mining Techniques

Clustering

- What are the ages, ethnicity, location, family size, and affluence of our clients?
- What are the ages, ethnicity, location, family size, and affluence of our clients by profit, number of products, and product groups?

Prediction

- Which prospects are most likely to buy a specific product or service?
- Which clients are at risk of defecting to competitors? What are the attributes associated with them leaving?

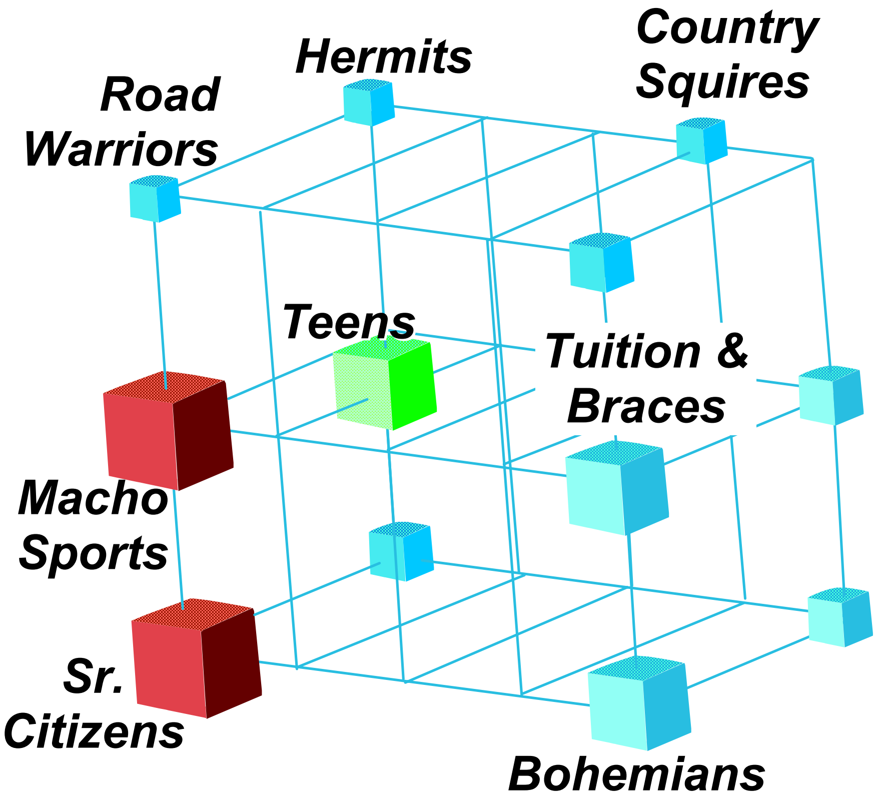
Affinity

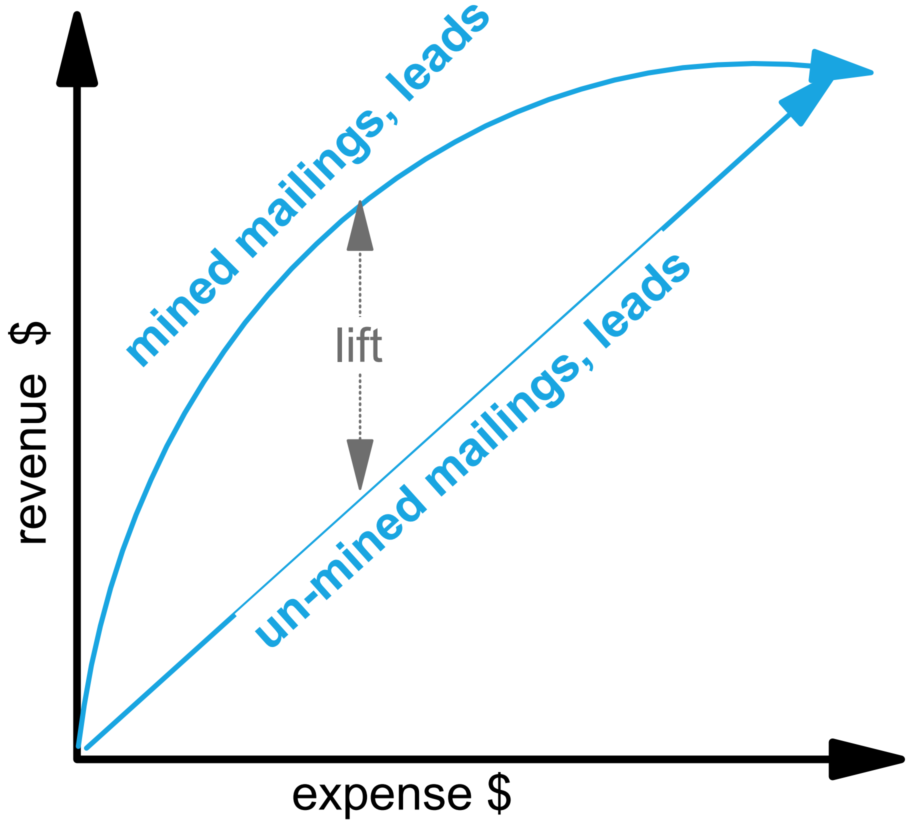
- Which products are most often purchased by the most profitable clients?
- Are there any product purchases that often trigger additional purchases later?

Sequences & Pattern Detection

- Are there any repeating patterns for customers who stop buying or cancel service?
- Where are there patterns of purchases and returns?

Typical Business Uses





Clustering & Segmentation

Prediction

DB2 Data Management Software

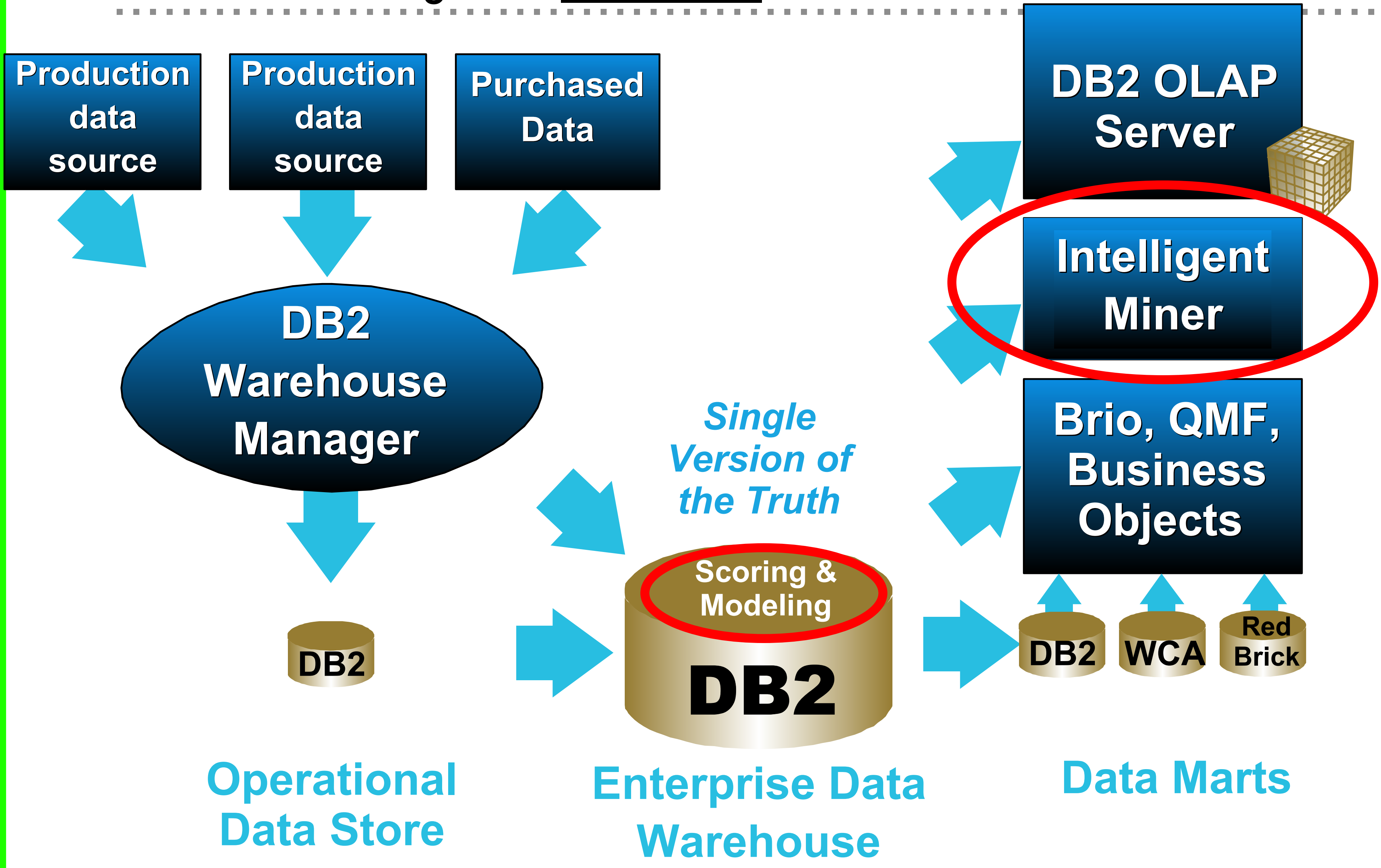
Clustering (banking example)

Segments	Segment Services used	Primary differentiator	Secondary
Bohemians	checking, credit card	high restaurant cc use	3X credit card usage
Road Warrior	Business credit line, checking, college funds, mortgage	business credit card & $5.5K monthly balance	out of country ATM use
Senior Citizens	checking, credit card	3+ savings accounts > $20K	mortgage principle < 18% loan
Country Squires	mortgage, credit line, checking, porfolio, platinum credit card	7X more likely to have boat loan	rental property
Tuition and Braces	mortgage, checking, credit card, college savings	Mortgage < 8 years old	2X credit card usage
• • •	• • •	• • •	• • •
Hermits	checking, credit card, mortgage	cc usage < 8 per month	electronics purchases

Prediction (banking example)

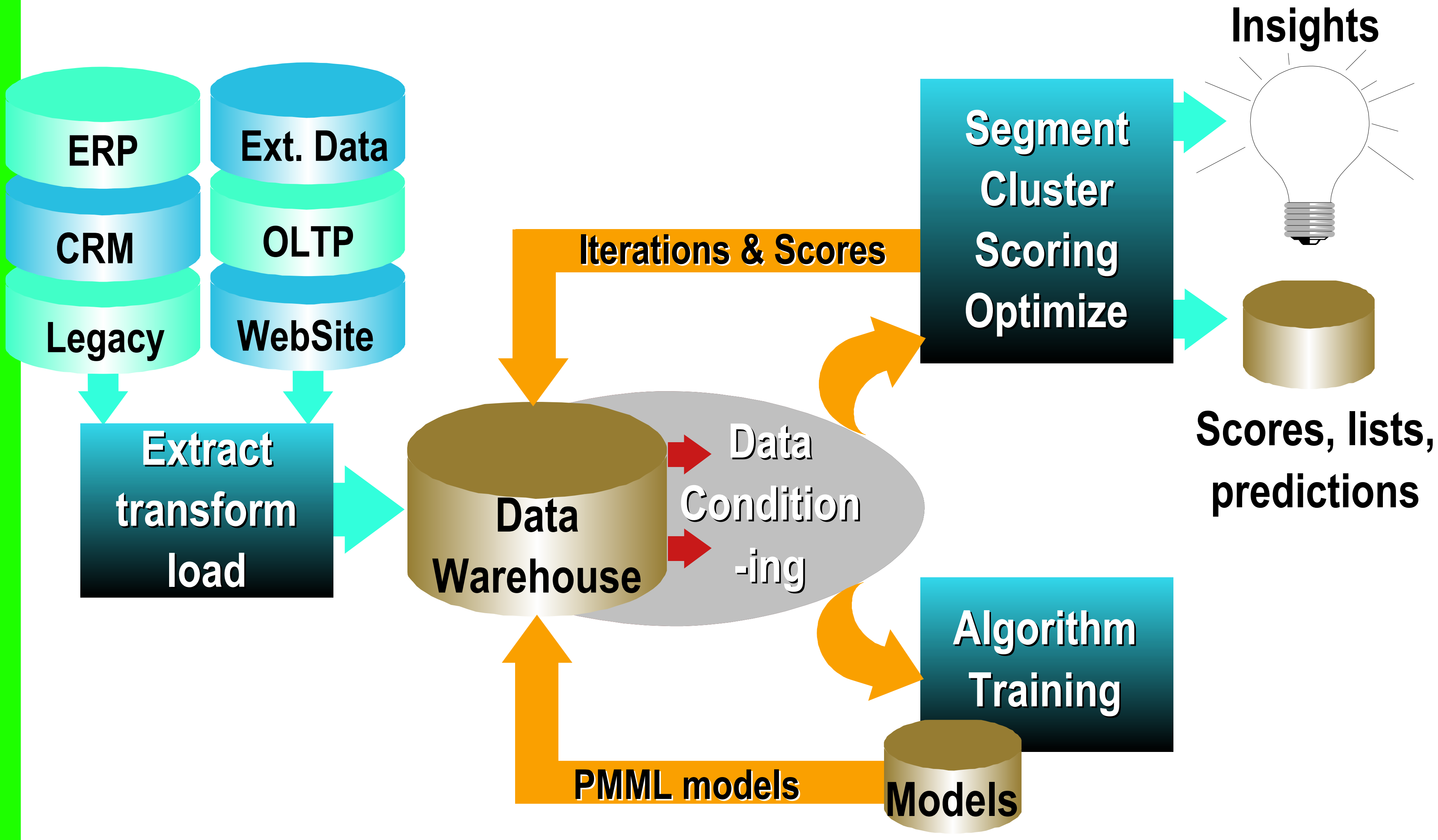
Segments	Will Buy	Will Attrite
Bohemians	▪ Merchant co-sponsored credit card 28% of time	▪ balance of 33 cc's reaches 85% of limit
Road Warrior	▪ Airline co-sponsored cc 43% of time ▪ 2nd home mortgage 11% of time	▪ business cc balance drops 70% for 3 consecutive months
Senior Citizens	▪ Bond fund ABC 21% of time ▪ Bond fund XYZ 17% of time	▪ 3 call center complaints < 4 months
Country Squires	▪ foreign country home mortgage 21% of time ▪ Bond fund XYZ 31% of time	▪ relocation & nearest branch office > 4 miles
Tuition and Braces	▪ Refi mortgage when 1.8% delta	▪ College fund < $10K and cc balance < $3K and call center complaints in last two months
● ● ●	● ● ●	● ● ●
Hermits		

Business Intelligence **Products**



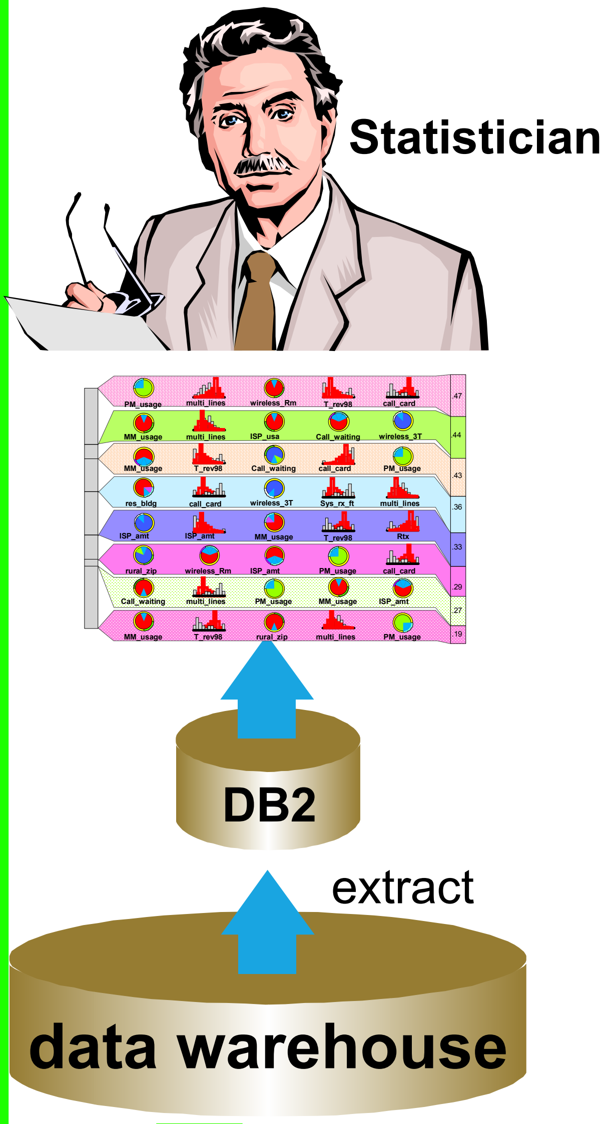
DB2 Data Management Software

The Data Mining process

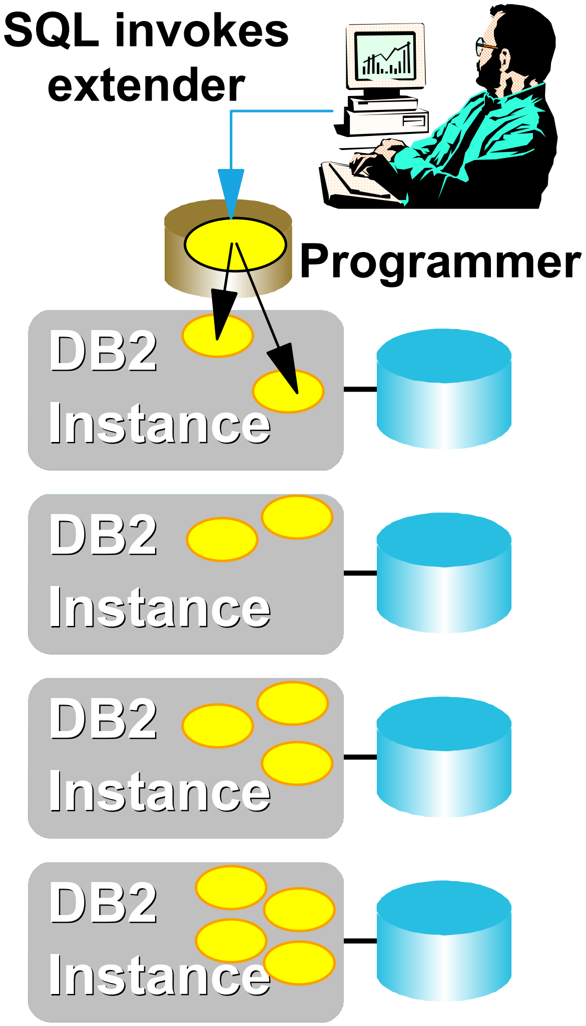


Data Mining products

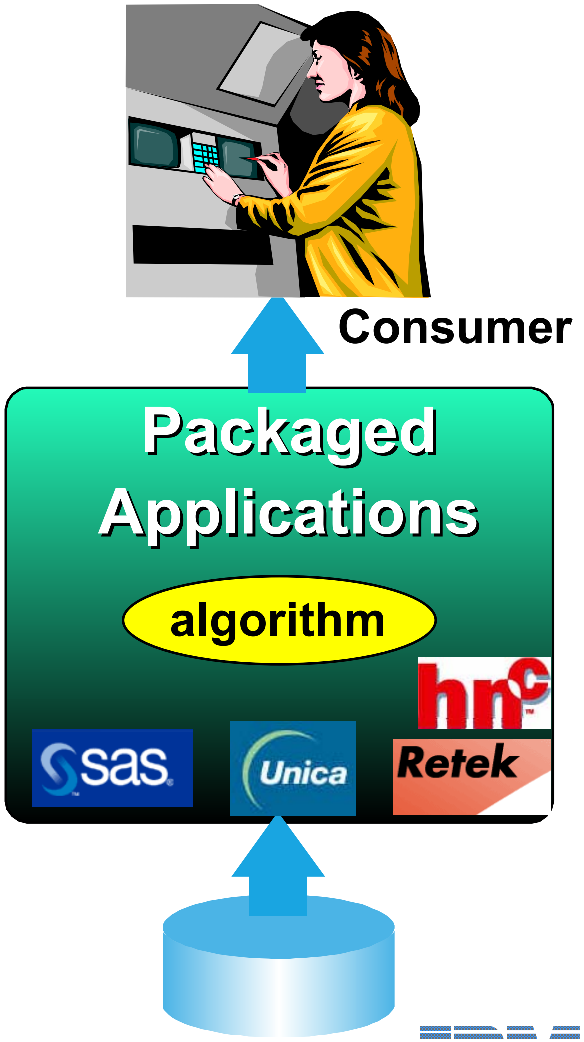
Workbench



RDBMS Extenders

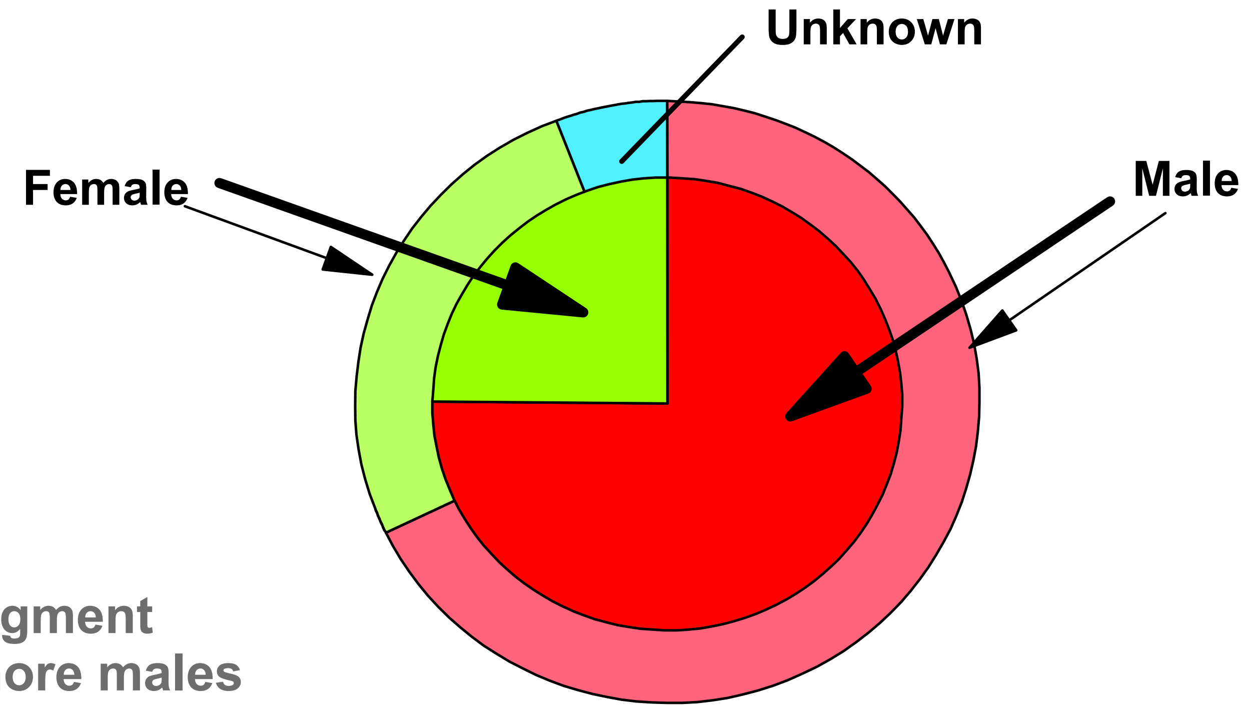


Application Embedded



Visualization of database attributes

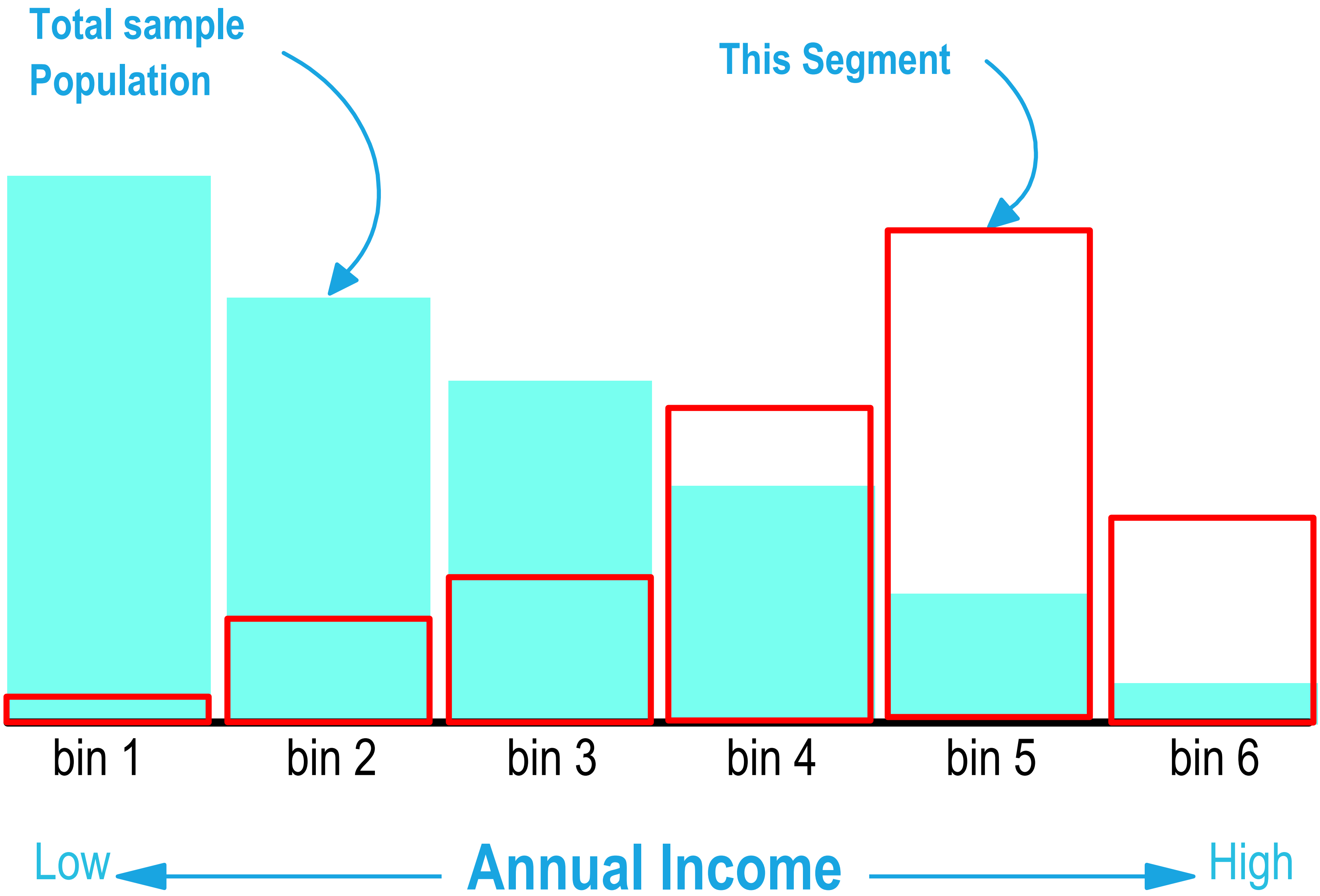
Categorical Variable
inner circle - this segment
outer circle - total population



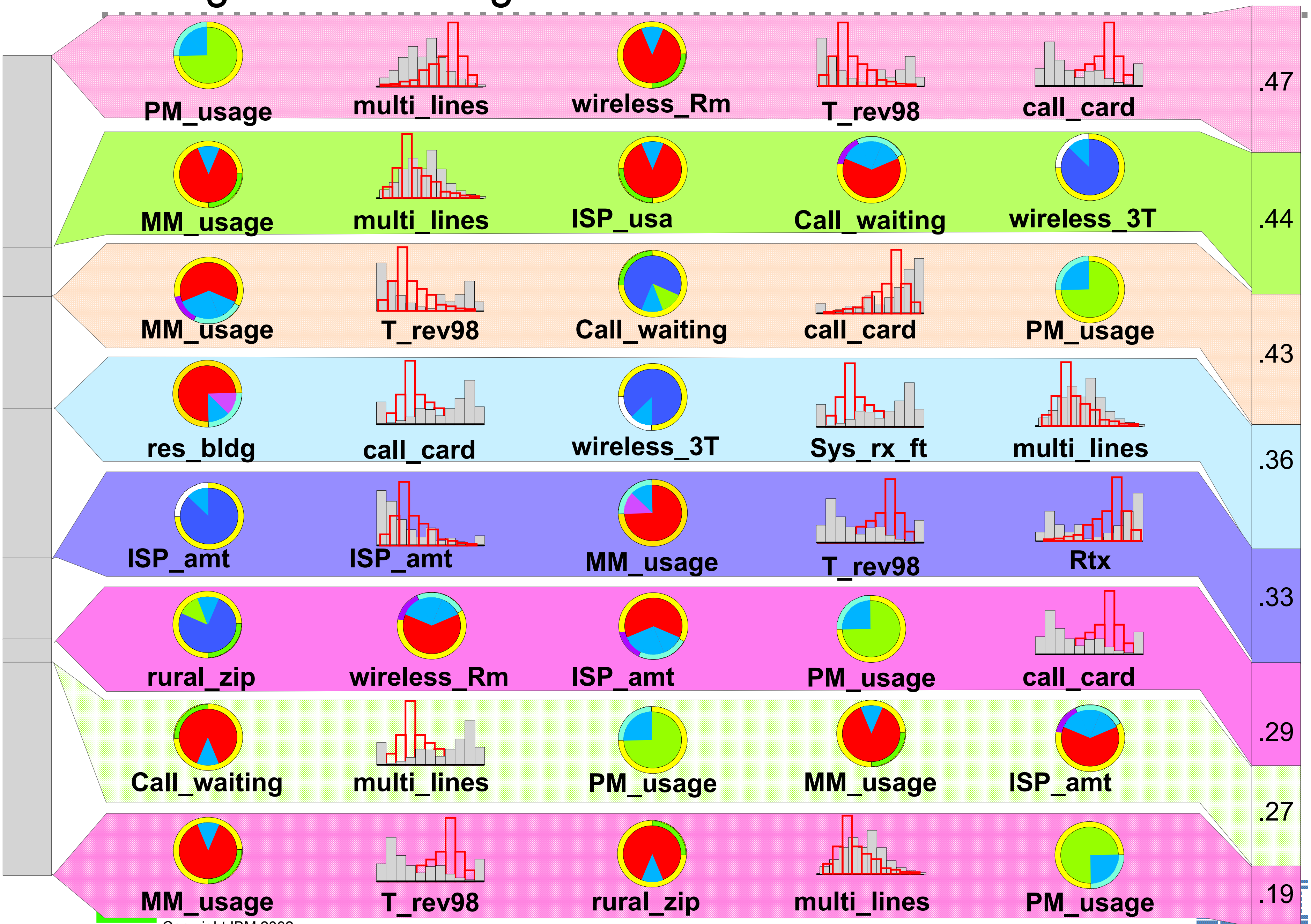
GENDER

This segment contains more males and fewer females than the overall population sample

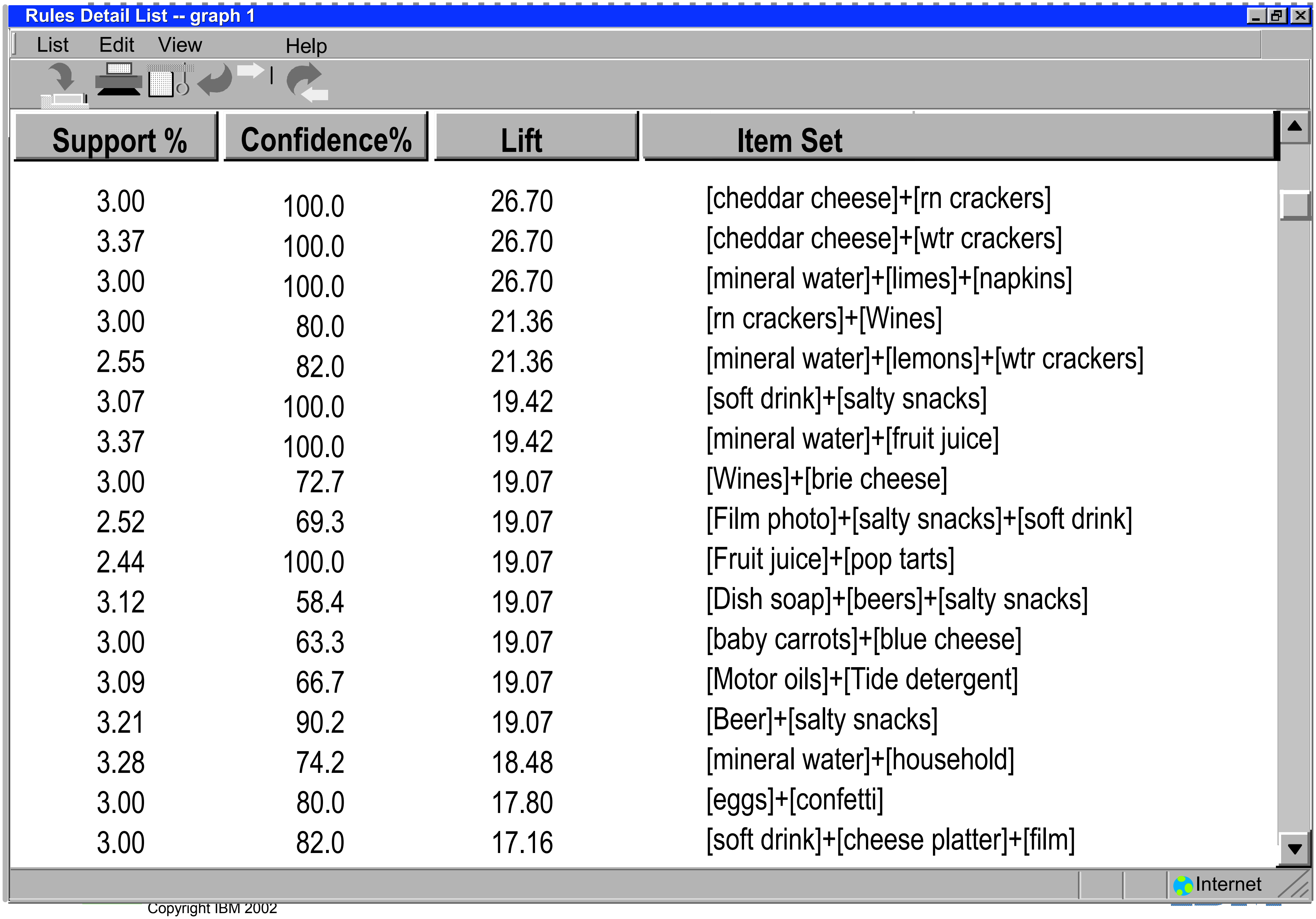
Continuous Variable Visualization



Intelligent Miner Segments Visualization



Associations Visualizer



Support %	Confidence%	Lift	Item Set
3.00	100.0	26.70	[cheddar cheese]+[rn crackers]
3.37	100.0	26.70	[cheddar cheese]+[wtr crackers]
3.00	100.0	26.70	[mineral water]+[limes]+[napkins]
3.00	80.0	21.36	[rn crackers]+[Wines]
2.55	82.0	21.36	[mineral water]+[lemons]+[wtr crackers]
3.07	100.0	19.42	[soft drink]+[salty snacks]
3.37	100.0	19.42	[mineral water]+[fruit juice]
3.00	72.7	19.07	[Wines]+[brie cheese]
2.52	69.3	19.07	[Film photo]+[salty snacks]+[soft drink]
2.44	100.0	19.07	[Fruit juice]+[pop tarts]
3.12	58.4	19.07	[Dish soap]+[beers]+[salty snacks]
3.00	63.3	19.07	[baby carrots]+[blue cheese]
3.09	66.7	19.07	[Motor oils]+[Tide detergent]
3.21	90.2	19.07	[Beer]+[salty snacks]
3.28	74.2	18.48	[mineral water]+[household]
3.00	80.0	17.80	[eggs]+[confetti]
3.00	82.0	17.16	[soft drink]+[cheese platter]+[film]

The PMML Interoperability Standard

Unica

SPSS

Epiphany

Startup
de Jure

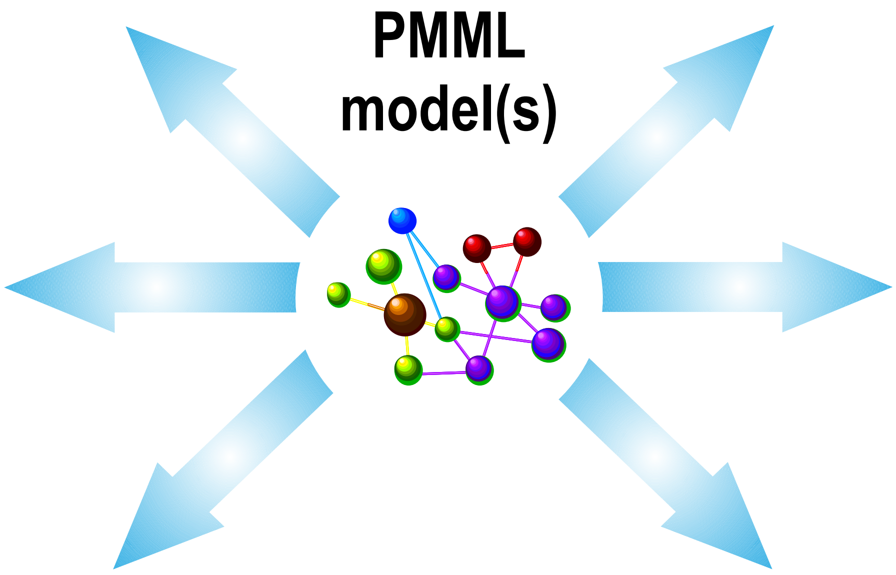
Walker

WebSphere
apps

Intelligent Miner

consume
pmml

emit
pmml



PMML
model(s)

emit
pmml

consume
pmml

SAS

SAP

Kana

Fair Isaac

Chordiant

Query Tools

WebTrends

Hyperion

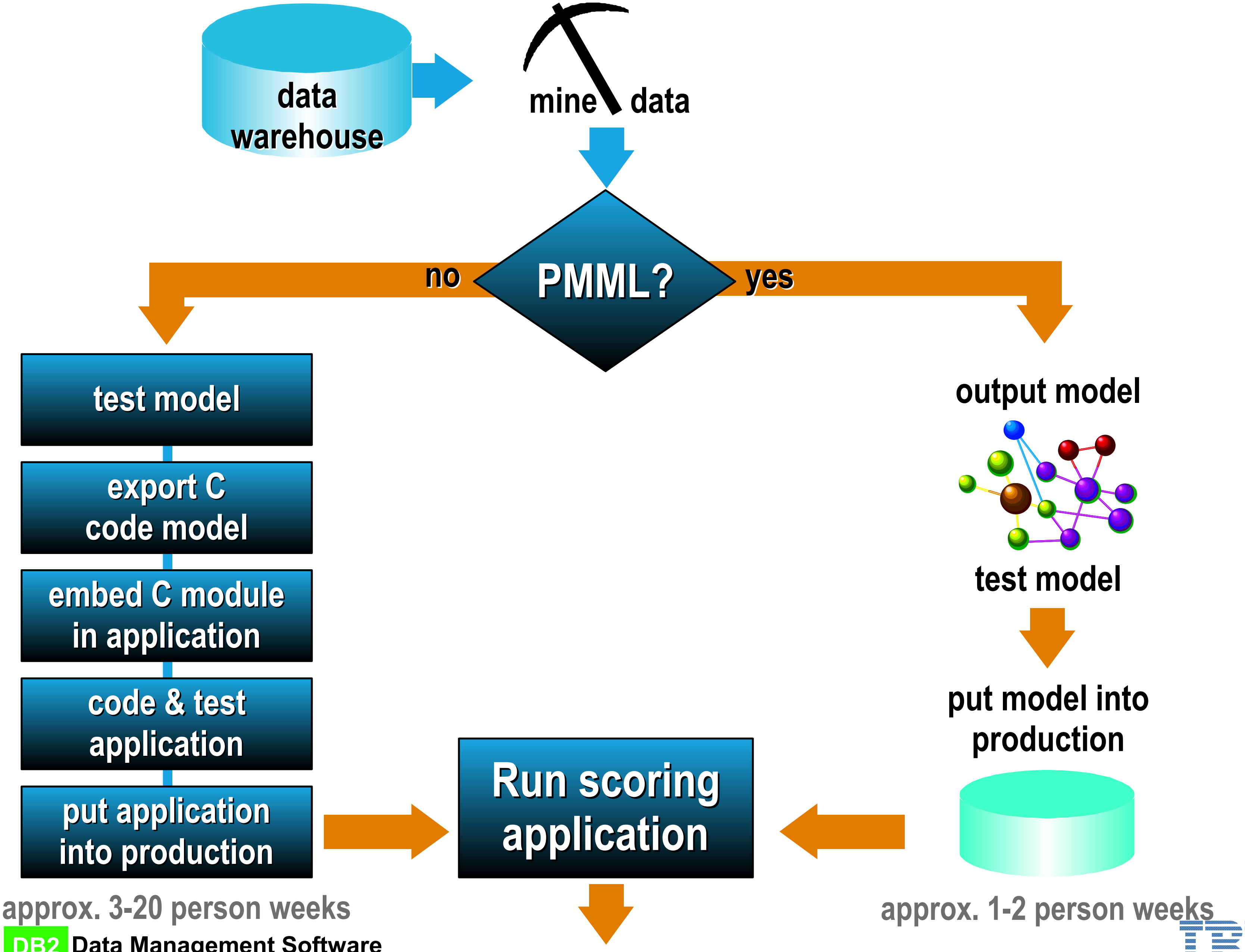
JDA

open standards enable innovation, partnerships, and choice

What is PMML?

- Predictive Model Markup Language
- PMML 2.0 is a standard for XML documents which express trained instances of analytic models. The following classes of model are addressed
  - Association Rules, Decision Trees , Center-Based & Distribution-Based Clustering , Polynomial Regression , General Regression, Neural Networks, Naive Bayes
- A standard for developing a data mining model -- the training set -- to be used by one or more consuming software tools
- PMML Standards supporters
  - Angoss Software, IBM, Magnify, Oracle, National Center for Data Mining University of Illinois, SAS, SPSS Inc., Xchange, MINEit Software, NCR, KXEN, ...
- http://www.dmg.org

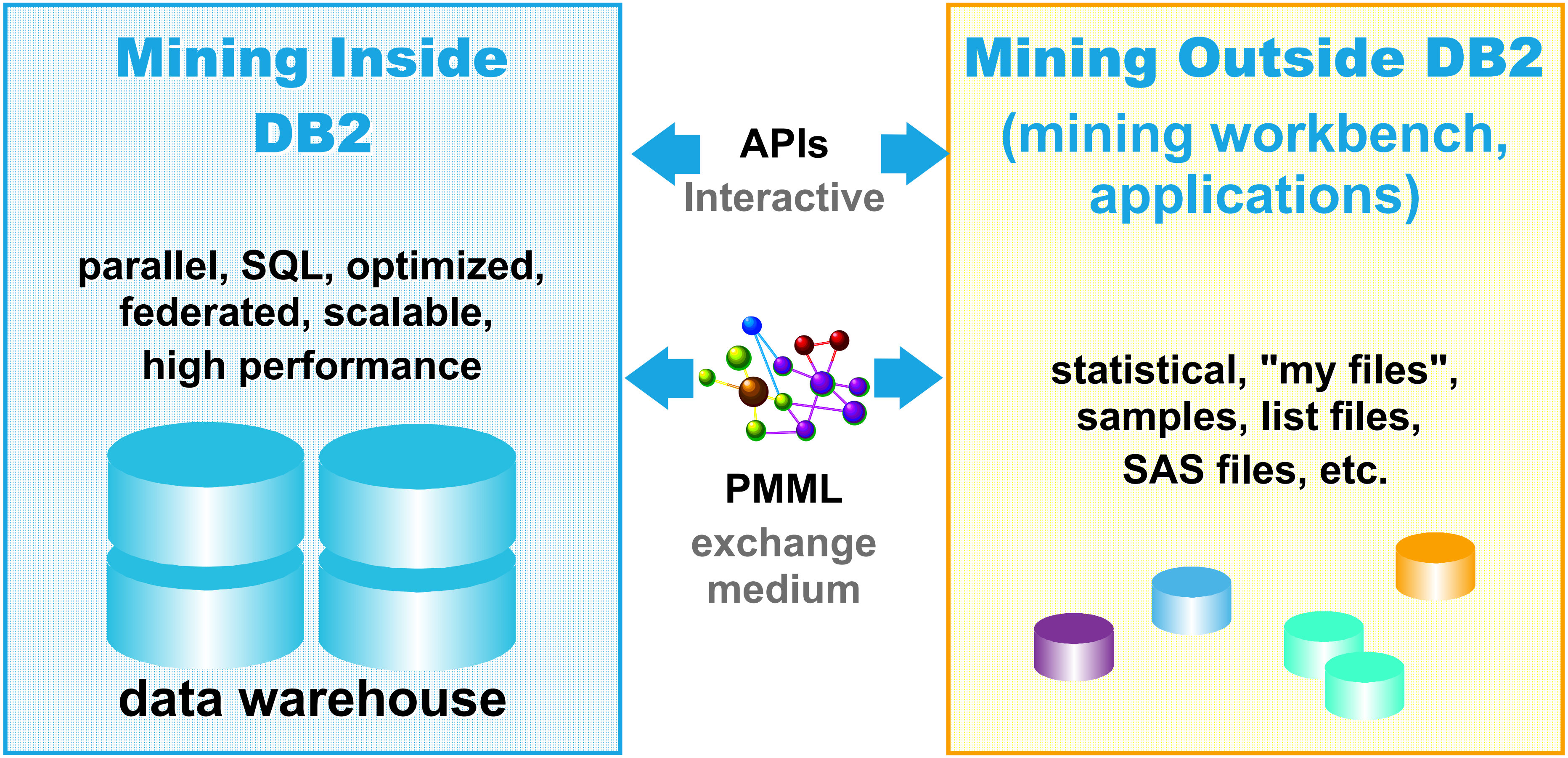
PMML Benefits - Reduced Labor costs



approx. 3-20 person weeks

approx. 1-2 person weeks

Mining Inside & Outside of the RDBMS



Mining Inside DB2

parallel, SQL, optimized, federated, scalable, high performance

data warehouse

APIs

Interactive

PMML

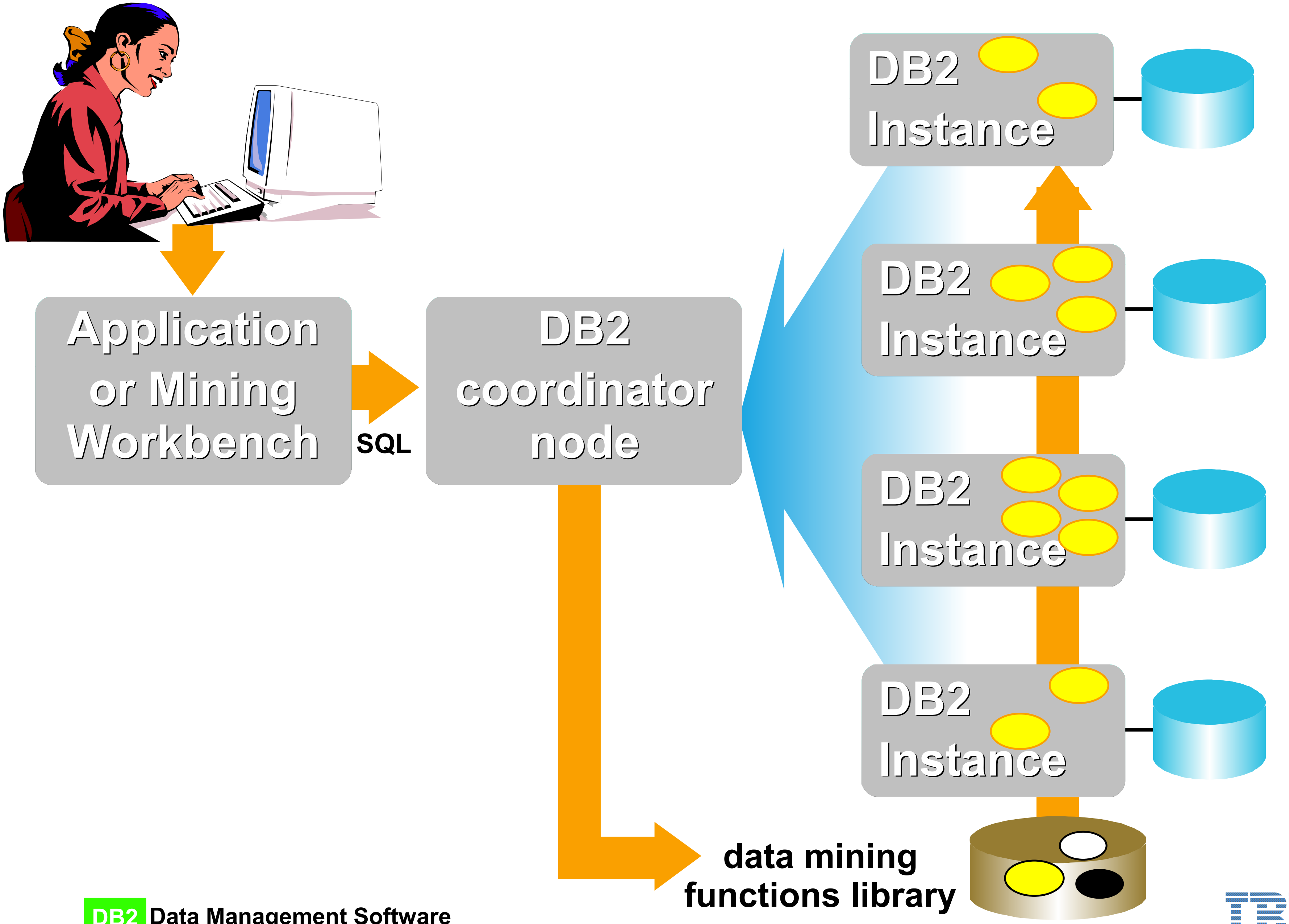
exchange medium

Mining Outside DB2

(mining workbench, applications)

statistical, "my files", samples, list files, SAS files, etc.

End User activates Mining Extenders



Application or Mining Workbench

SQL

DB2 coordinator node

DB2 Instance

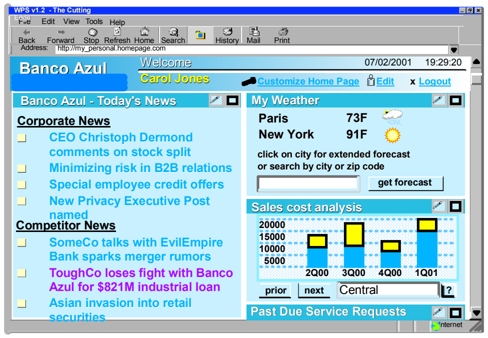
DB2 Instance

DB2 Instance

DB2 Instance

data mining functions library

Applications Connect to Real Time Data Mining



WPS v1.2 - The Cutting Edge

File Edit View Tools Help

Back Forward Stop Refresh Home Search History Mail Print

Address: http://my_personal.homepage.com

Banco Azul

Welcome 07/02/2001 19:29:20

Carol Jones

Customize Home Page | Edit | x Logout

Banco Azul - Today's News

Corporate News

- CEO Christoph Dermond comments on stock split
- Minimizing risk in B2B relations
- Special employee credit offers
- New Privacy Executive Post named

Competitor News

- SomeCo talks with EvilEmpire Bank sparks merger rumors
- ToughCo loses fight with Banco Azul for $821M industrial loan
- Asian invasion into retail securities

My Weather

Paris	73F
New York	91F

click on city for extended forecast or search by city or zip code

get forecast

Sales cost analysis

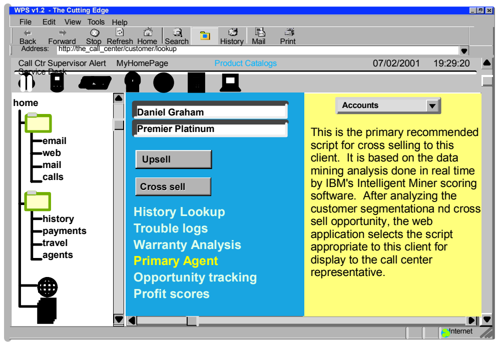
prior | next | Central | ?

Past Due Service Requests

Internet

BI Reports

Portal KPIs



WPS v1.2 - The Cutting Edge

File Edit View Tools Help

Back Forward Stop Refresh Home Search History Mail Print

Address: http://the_call_center/customer/lookup

Call Ctr Supervisor Alert MyHomePage Product Catalogs 07/02/2001 19:29:20

Service Desk

home

- email
- web
- mail
- calls

- history
- payments
- travel
- agents

Daniel Graham

Premier Platinum

Upsell

Cross sell

History Lookup

Trouble logs

Warranty Analysis

Primary Agent

Opportunity tracking

Profit scores

Accounts

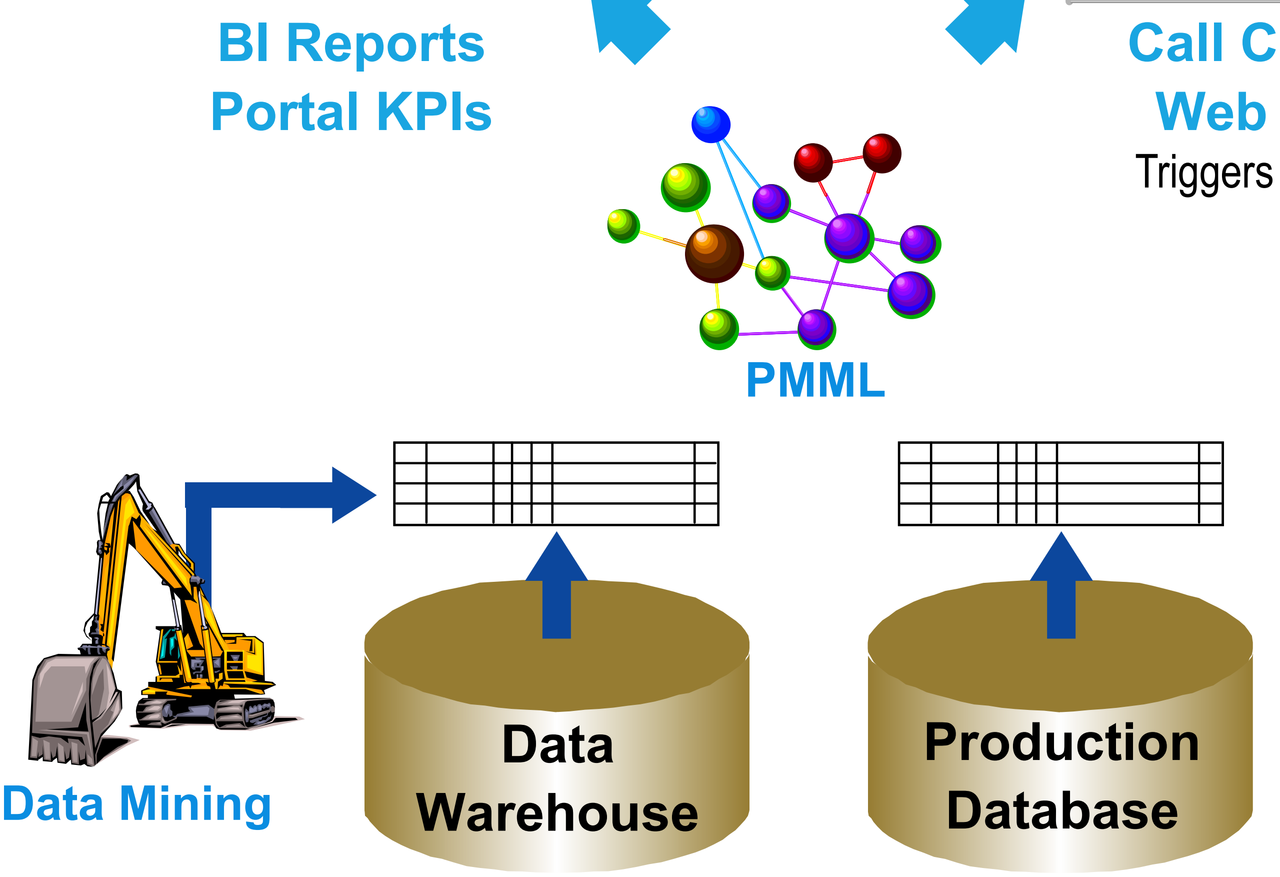
This is the primary recommended script for cross selling to this client. It is based on the data mining analysis done in real time by IBM's Intelligent Miner scoring software. After analyzing the customer segmentationa nd cross sell opportunity, the web application selects the script appropriate to this client for display to the call center representative.

Internet

Call Center

Web Site

Triggers or Agents



PMML

Data Mining

Data Warehouse

Production Database

DB2 Data Management Software

Benefits of DB2 Extenders Data Mining

- **Data mining can be performed using parallel processing**
  - Workload scalability --big tasks
  - SMP and/or cluster parallelism = response time reduction
    - More "runs" per day for deeper accuracy and more discoveries
- **Avoiding extracts & loads -- less data movement & redundancy**
  - less labor costs, less processing = lower costs overall
- **EDW Server has the bigger faster CPUs, memory, disks, etc.**
  - CPU intensive workloads finish quicker
- **More data can be mined resulting in more detailed analysis**
  - Sometimes "sample" subsets aren't enough for accuracy
    - e.g. deviation detection
- **Scoring Extenders can be leveraged in online applications**

SQL Modeling, example

```
call IDMMX.BuildClusModel(
          'CustomerSegments',
          IDMMX.DM_MiningData()
          ..DM_defMiningData('BANKING_CUST'),
          IDMMX.genClusSettings(
                  IDMMX.DM_LogicalDataSpec()
                  ..DM_genLogicalDataSpec('BANKING_CUST')
                  ..DM_remDataSpecFld( 'GENDER' ) )
          ..DM_setMaxNumClus(9)
          ..DM_setFldUsageType( 'PRODUCT', 2 )
          ..expClusSettings()   );
```

SQL Scoring, example

```
SELECT
   d.name, d.age,
   IDMMX.DM_getClusterID(
         IDMMX.DM_applyClusModel(
              cm.model,
              IDMMX.DM_impApplData(
                  REC2XML(1,'COLATTVAL','',
                         d.age, d.salary,
                         d.region,d.product,
                         ..., d.goldcard)))
   )
FROM ClusterModels cm, MyData d
WHERE cm.modelname='CustomerSegments';
```

SQL mining 'macros', ease of use

```
Call CreateClassifier(
        'Campaign',      -- name of new classification model
        'Customer',      -- input data
        'Response' );    -- target field


Call CreateScoringView(
      'CustomerWithScore',   -- name of new SQL VIEW
      'Customer'             -- source table
      'Campaign');           -- prediction model
```

SQL mining 'macros', ease of use

```
Select ColumnName, Rank
From Table ( IDMMX.InfluenceFactors(
                 'CustomerView', 'SalesPerMonth') ) F
order by rank descending;
```

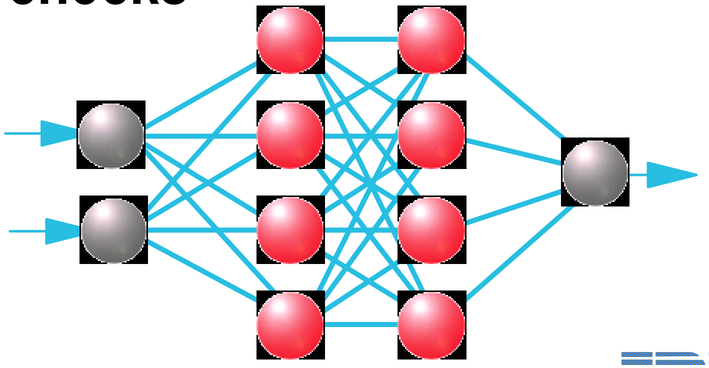
-- returns table of colunms in 'CustomerView'

-- ranked by how much they are related to 'SalesPerMonth'

```
  ColumName,    Rank
--------------------------------
   Location        0.64
   Age             0.43
   Gender          0.142
   ...             ...
```

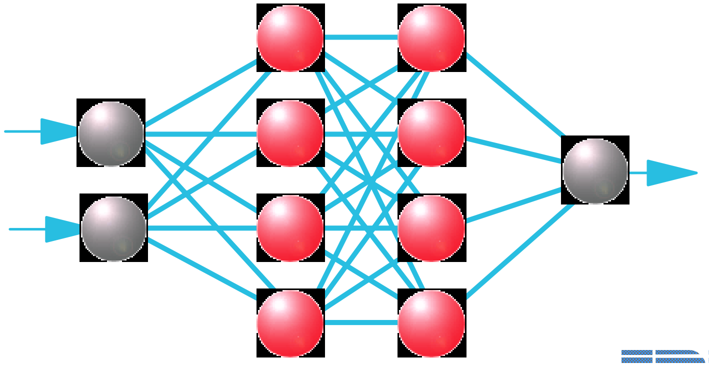
Trends, directions for integrated mining

- **Ease of use, vertical integration**
  - ► **Mining integrated into domain specific applications**
    - **E.g. prediction in campaign management**
  - ► **End-user does not (have to) know mining**
  - ► **Mining expertise encapsulated in app.design**
  - ► **Robust, smart mining algorithms, visualization components**
- **Ease of use, intelligent helpers**
  - ► **Find important influence factors**
  - ► **Outliers, anomalies, data quality checks**
  - ► **Guided OLAP, 'drill mine'**



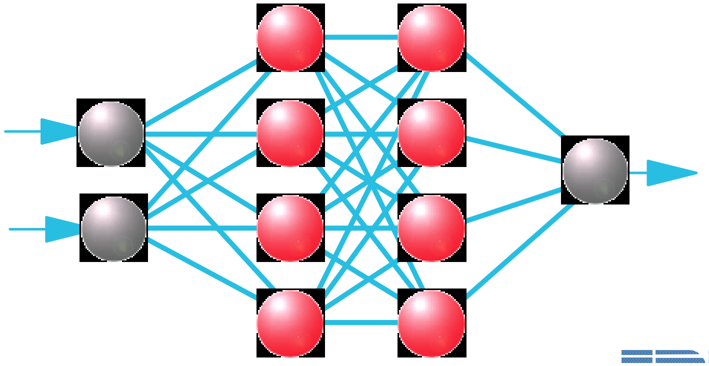
Trends, directions for mining in the database

- High performance, scalable
- Simple maintenance
  - common API, no separate mining tool admin
- Ease of use, intelligent helpers
  - 'just another' simple database function



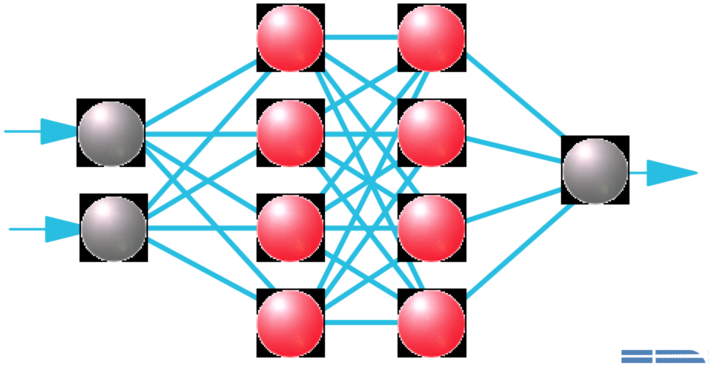
Trends, directions for standards

- PMML
  - Model exchange format
    - e.g., SAS Modeling -> DB2 Scoring
  - Challenge: 'my algorithm is better than yours'
- SQL: ISO SQL/MM, Microsoft OLE DB
  - "2 + 1" major players
- Java
  - New standard, extensive API
- Web Services, XMLA
  - Business uses? Scoring.



Trends, directions for mining

- End of mining workbenches for experts?
  - No, still needed for flexible data analysis
  - But, also some support for 'mining' in other BI tools.
- Database API with mining
  - For heavy-lifting, back-end to mining tools & apps
  - Integrated with other transformations & information flow
  - Easy to use, intelligent helper functions that are packaged with the data warehouse server at no additional cost.



Summary

- **Data Mining is vital to most industries, especially in CRM-Analytics**
- **Businesses use Data Mining for**
  - Lead generation & mailing lists
  - Cross-sell and Up-sell predictions
  - Anomaly detection
- **Intelligent Miner For Data & Extenders:**
  - Run Inside DB2 --fast and in parallel
  - Have the most algorithms to fit the most business needs
  - Setting the pace with Standards
- **Easy mining**