TOC

INDEX

VIEW

B22 Data mining in action -- Real-time Scoring

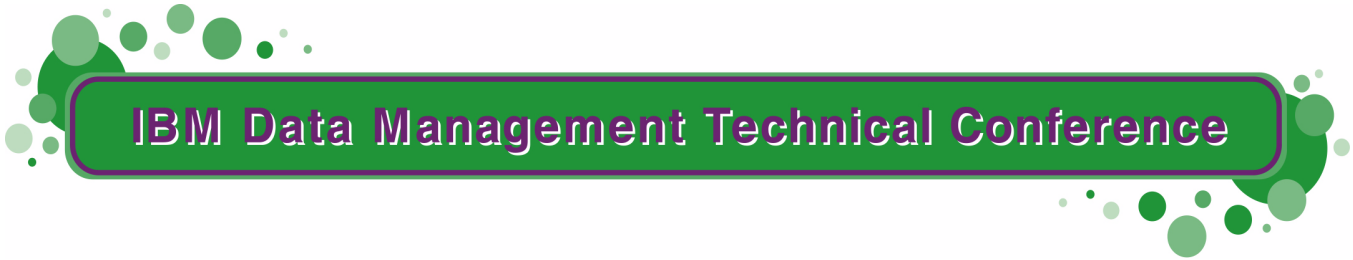
Gregor Meyer, Senior Software Engineer, IBM

The access to predictive mining results closes the loop between operational applications and data warehouse analytics. For example, a call center application can automatically enrich the current customer information with a predicted churn risk. The prediction is computed by a data mining scoring function. The presentation will show a series of examples how DB2 Intelligent Miner Scoring is used in customized applications or in partner tools such as Business Objects or SAP. The scoring functions are simple standard extensions to SQL. They can be combined with any SQL query, VIEW, or TRIGGER. In certain e-business scenarios the input data for scoring may include data which is not yet made persistent in the database. Think of personalization in WebSphere where the current data might depend on the most recent mouse-click on a web page. A small Java API for the scoring functions allows for high speed predictive mining in these cases as well.

B22

Data mining in action
Real-time Scoring

Gregor Meyer



Anaheim, CA **Sept 9 - 13, 2002**

™!IBM Corporation 2002

RETURN TO INDEX

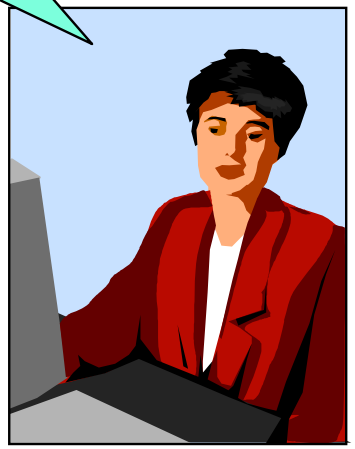
Agenda

- What is real-time scoring
  - Scenario
  - Predictive mining
- BI architecture
  - Data warehousing
  - "close the loop"
  - Real-time BI
- Data mining
  - Machine learning algorithms
  - Mining rules
- Scoring functions
  - SQL
  - Java

™!IBM Corporation 2002

Scenario

We have a special offer for you. Would you like to receive our weekly InvestmentGuide for letter for just $20 per quarter?



Business Intelligence Applications

™!IBM Corporation 2002

Data Mining

Data Mining is the process of extracting previously unknown, comprehensible, and actionable information from large databases.

™!IBM Corporation 2002

Data Mining

Data mining is about prediction

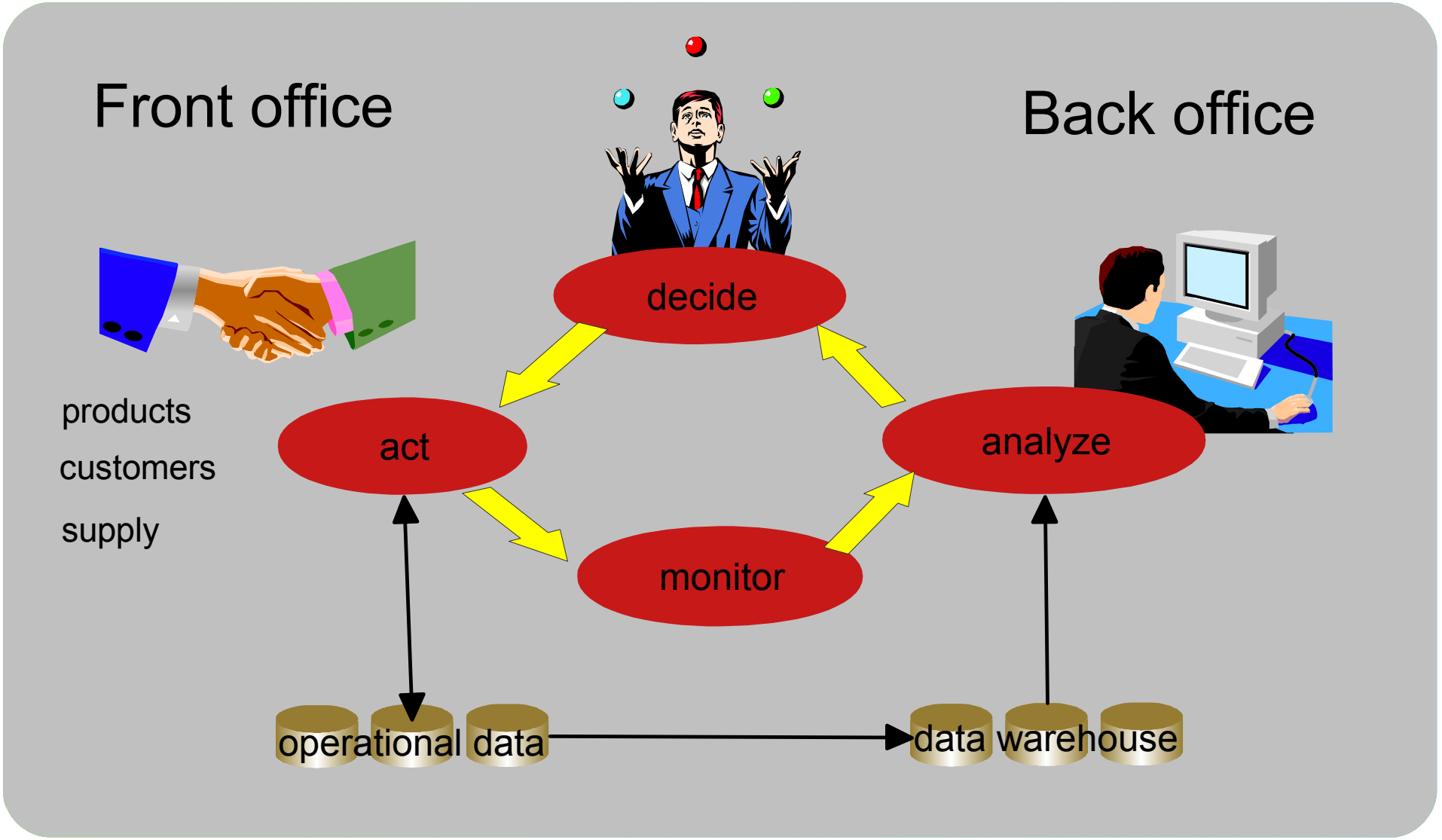
™!IBM Corporation 2002

Data mining examples

- Predict customer behavior
  - optimize mailing campaign
  - predict churn risk
  - highlight deviations
  - recommend products
  - predict demand for products
- Get Explanations
  - customer is likely to churn because he made many long distance phone calls during the night and many other customers with these characteristics went away.
  - Entry deviates because sales for product A in store B is much lower than expected
- Discover implicit rules and patterns
  - Discover that a certain kind of cheese sells with expensive wine
  - Discover a small group of customers with a similar behavior that makes them very profitable
- Act on results
  - Define co-sell and up-sell strategies, change pricing.
  - Contact customer who is likely to churn
  - Automatically recommend a product to a customer at the web site.

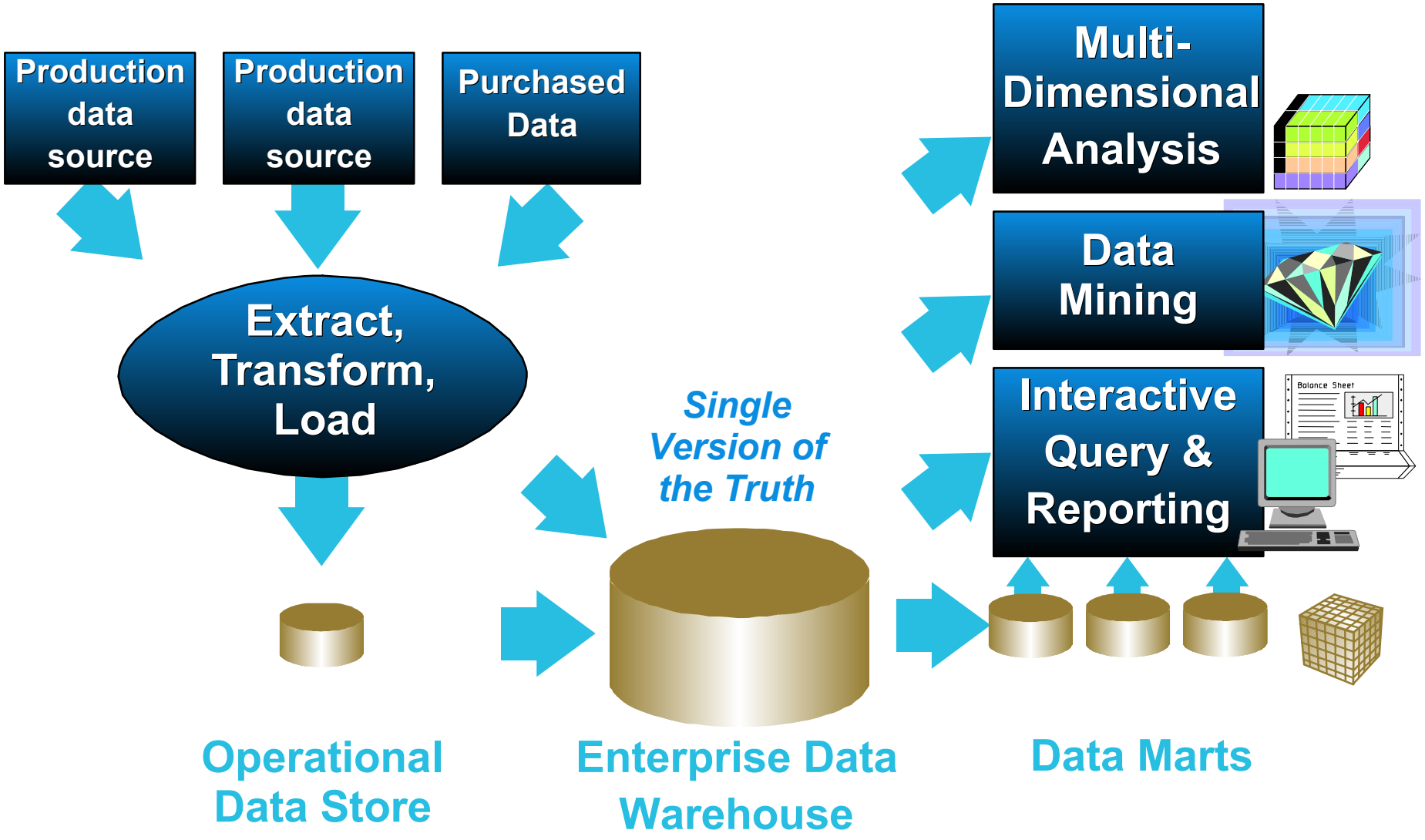
™!IBM Corporation 2002

BI, closed loop



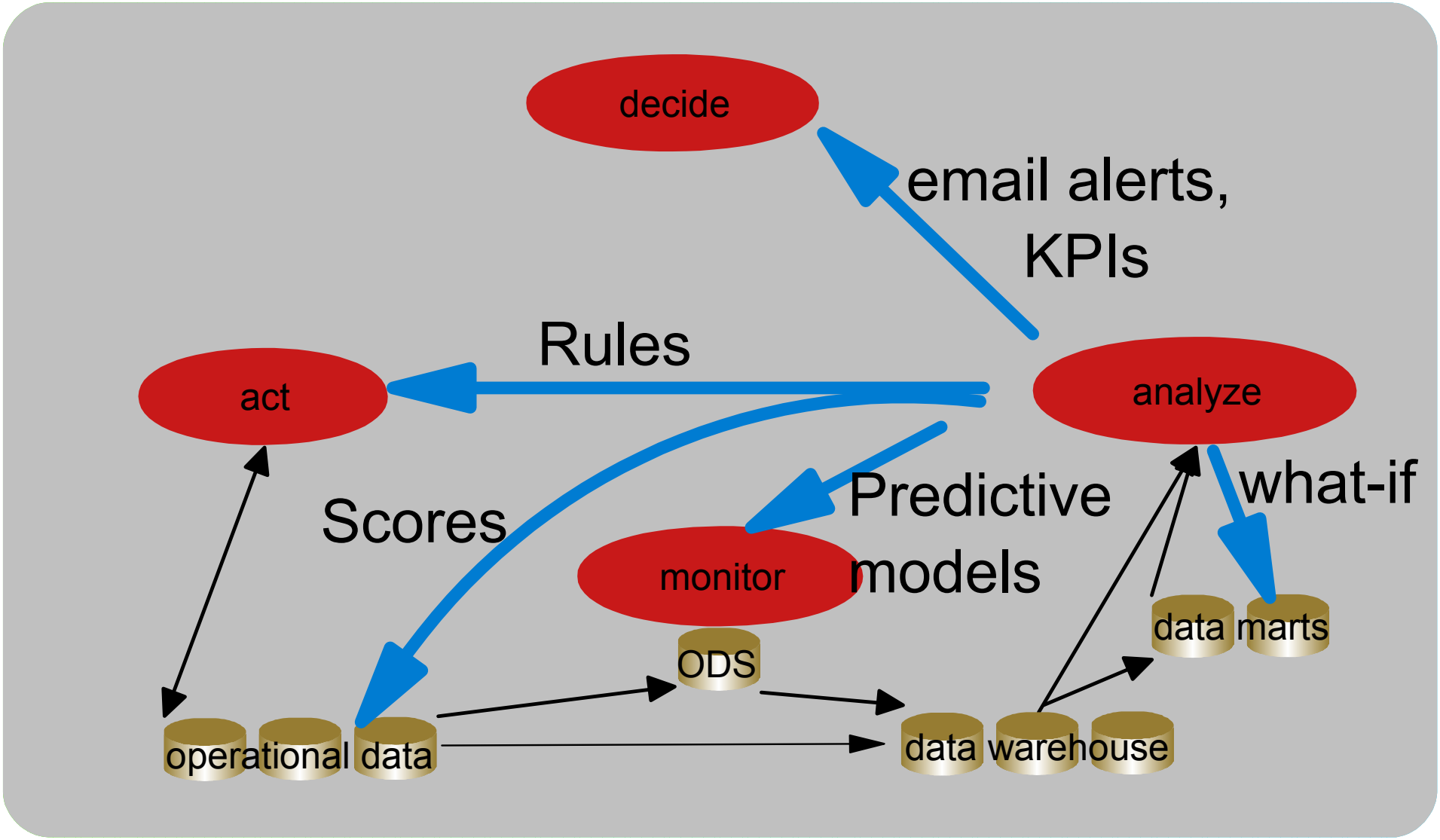
™!IBM Corporation 2002

Business Intelligence Infrastructure



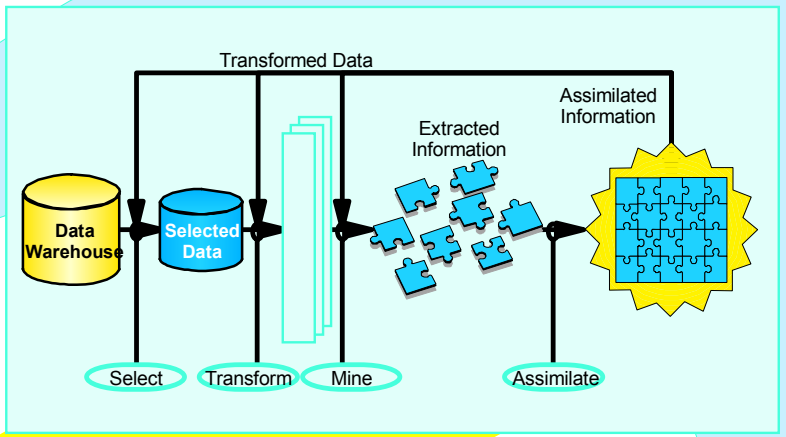
™!IBM Corporation 2002

BI, closed loop /2



DB2 Intelligent Miner

- Define the Problem
- Scope the Project
- Identify Data Sources
- Form the Team



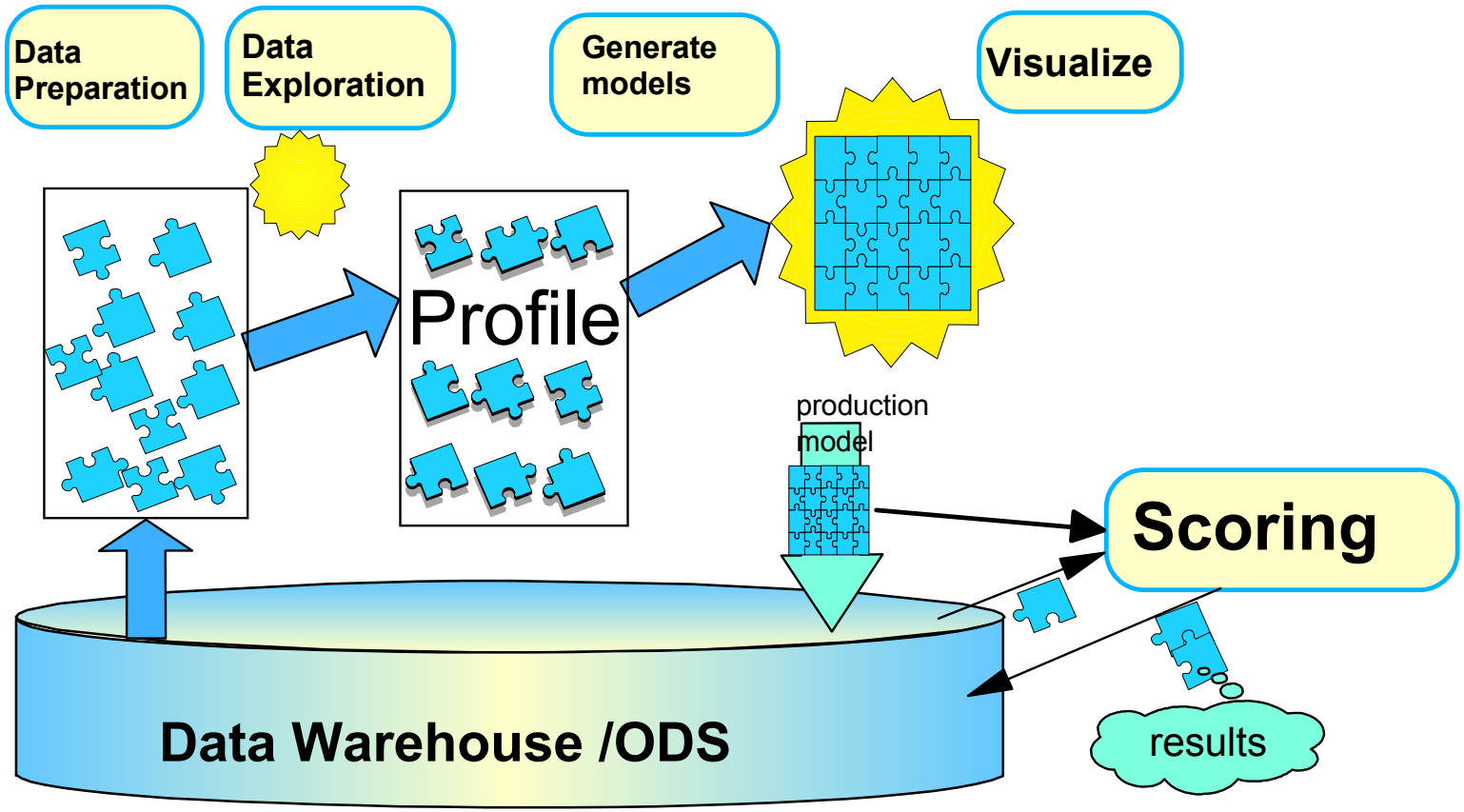
- Take Action
- Measure Results
- Assess Permanent Adoption

Data Mining
- *An exploratory process*
- *An experienced analyst*
- ➡ ***DB2 Intelligent Miner***

Deployment
- *A repeatable process*
- *Real-time prediction*
- ➡ ***DB2 Intelligent Miner Scoring***

™!IBM Corporation 2002

The Data Mining Process



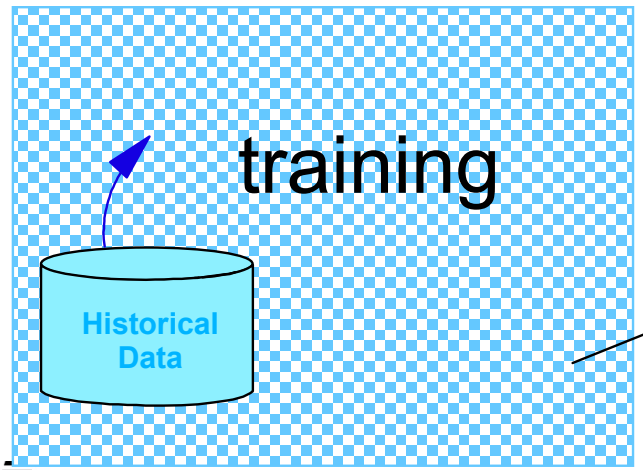
Deployment to Operational Systems

Customers / Suppliers

Business Users

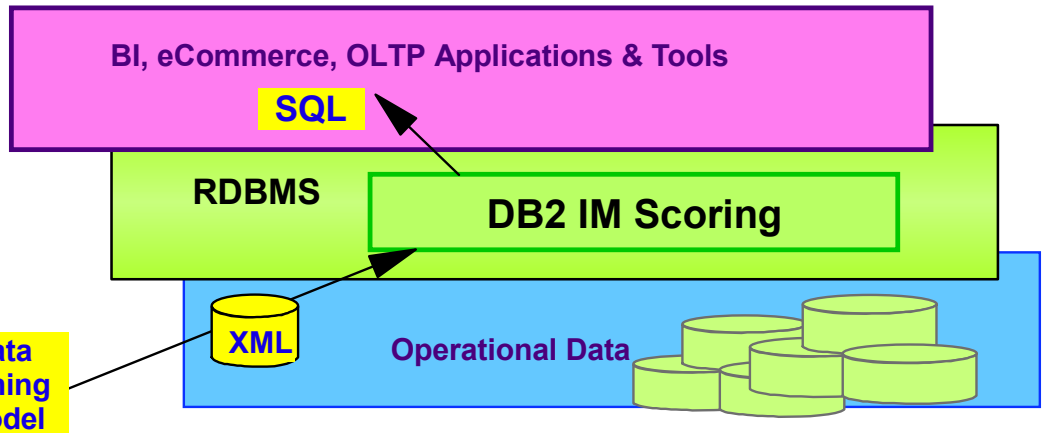
- Risk Assessment
- Fraud and Abuse Detection
- Promotion Targeting
- Best Practices
- Process Optimization

Data Mining Analyst



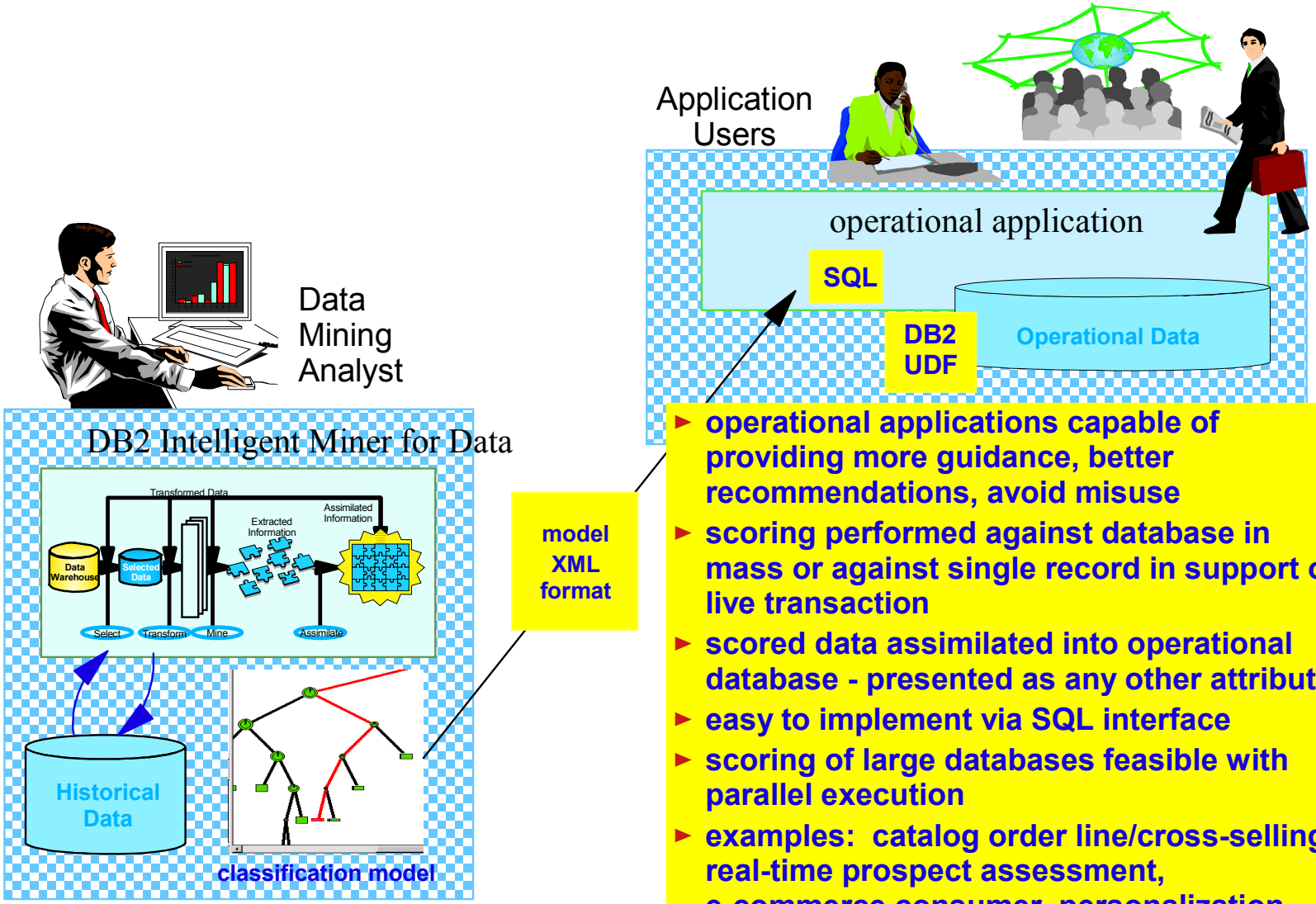
data mining model

PMML



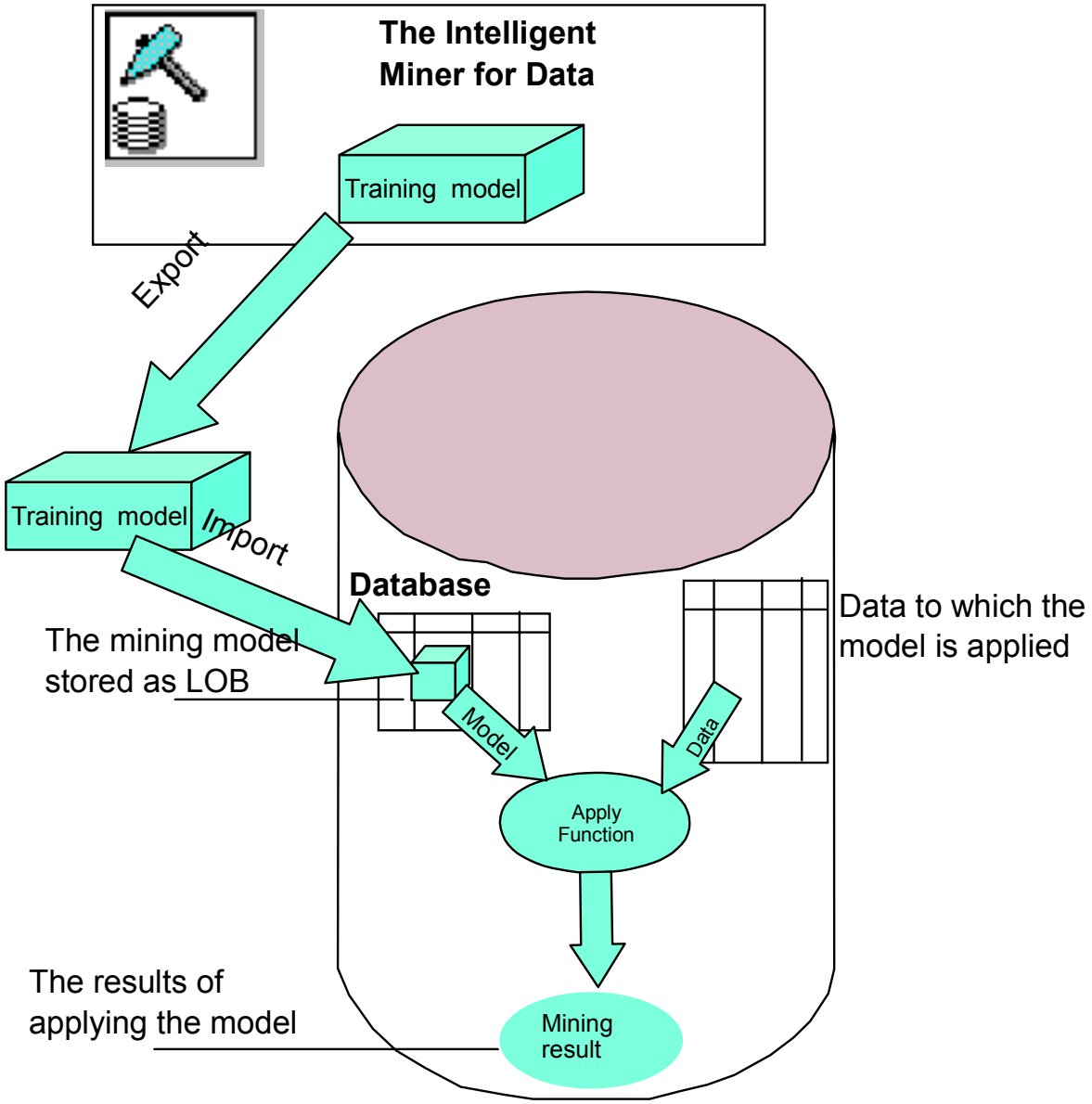
™!IBM Corporation 2002

Scoring Operational Data



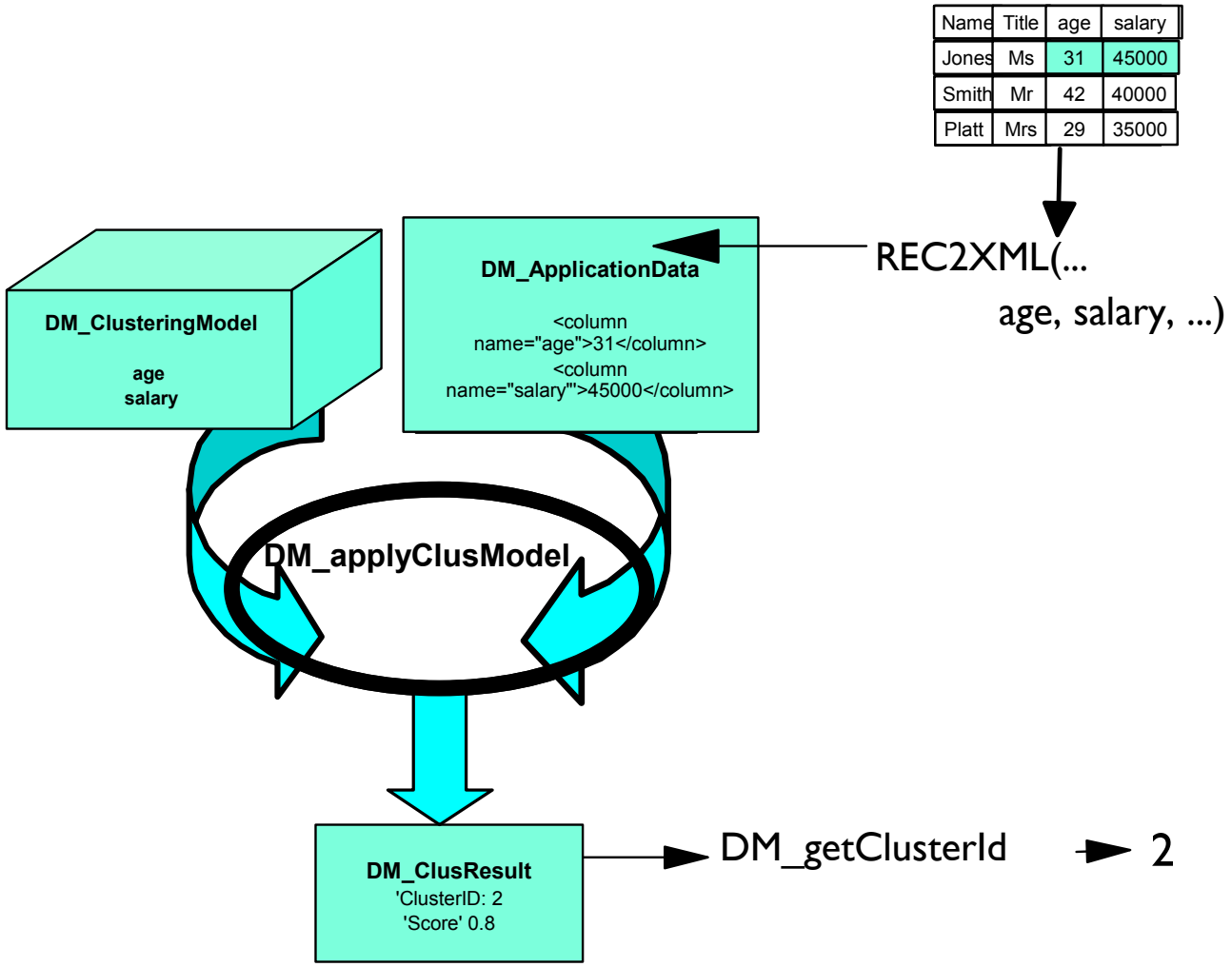
- operational applications capable of providing more guidance, better recommendations, avoid misuse
- scoring performed against database in mass or against single record in support of live transaction
- scored data assimilated into operational database - presented as any other attribute
- easy to implement via SQL interface
- scoring of large databases feasible with parallel execution
- examples: catalog order line/cross-selling, real-time prospect assessment, e-commerce consumer personalization

Scoring Process



™!IBM Corporation 2002

Applying a model



IM Scoring components

- SQL extension for data mining
- Java API: functionality same as in SQL
- PMML 2.0
  - Predictive Modeling Markup Language
  - mining models in XML
  - exchange models between vendors
  - www.dmg.org: IBM, SPSS, SAS, NCR, Microsoft, Oracle, Angoss, Xchange, SIGKDD,...

™!IBM Corporation 2002

Example SQL

```
INSERT INTO ClassifModels values
( 'DemoBanking',
IDMMX.DM_impClasFile('/tmp/demoBanking.mdl');

SELECT
  d.name, d.age,
  IDMMX.DM_getPredClass(
   IDMMX.DM_applyClasModel( cm.model,
    IDMMX.DM_impApplData(
     REC2XML(1,'COLATTVAL','',
             d.age, d.salary,d.gender)))
FROM  ClassifModels cm, MyData d
WHERE cm.modelname='DemoBanking';
```


™!IBM Corporation 2002

Real-time scoring Views

```
-- one-time setup
CREATE VIEW CustomerScore(
  CustId int,
  CreditRisk Double)
AS ( Select ... DM_getConfidence .. )


-- real-time Scoring in plain SQL

SELECT CustId, CreditRisk
FROM CustomerScore
WHERE CustId=1093
```


™!IBM Corporation 2002

DB2 Intelligent Miner Scoring

- ✓ Reuse of DB2 Intelligent Miner for Data mining technology delivers proven, industry-leading mining technology to operational applications
- ✓ Use of DB2 ensures scalability, reliability, federated data access
- ✓ DB2 extender implementation isolates the application from model interpretation and soring logic
- ✓ SQL API simplifies implementation, lowers skill requires to deploy mining across the enterprise
- ✓ Java API for in-memory scoring
- ✓ Managing models in the database ensures model integrity, ease of maintenance, timely model updates
- ✓ SQL WHERE clause makes it just as easy to score 1 case in real time, or all rows in mass
- ✓ Implementation to PMML standards offers an integration point for the use of a variety of modeling and CRM offerings
- ✓ Oracle cartridge implementation is also available

IBM®

™!IBM Corporation 2002