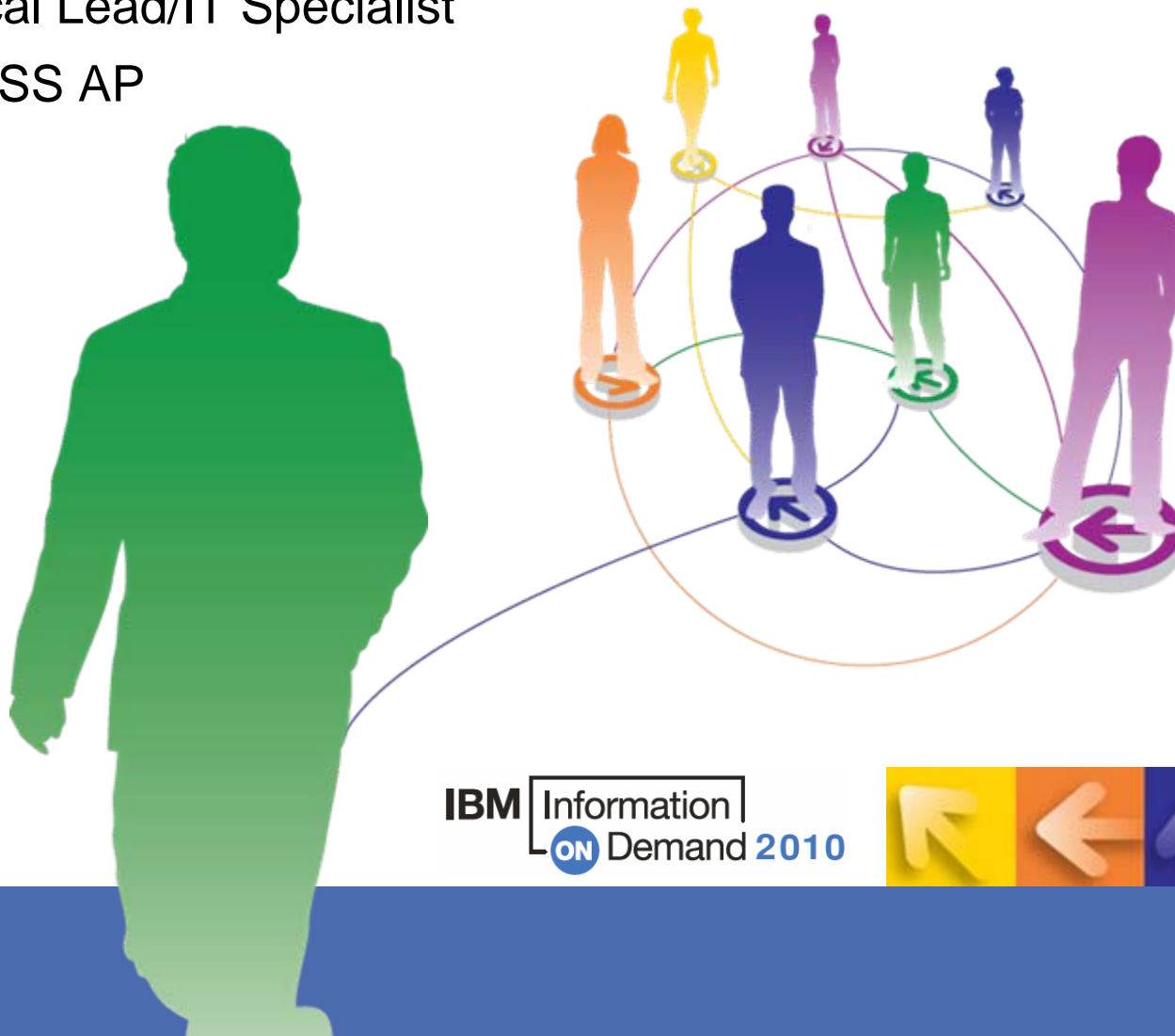


SPSS Predictive Analytics

A Technology View to Tomorrow



Alex Lee
Technical Lead/IT Specialist
IBM SPSS AP



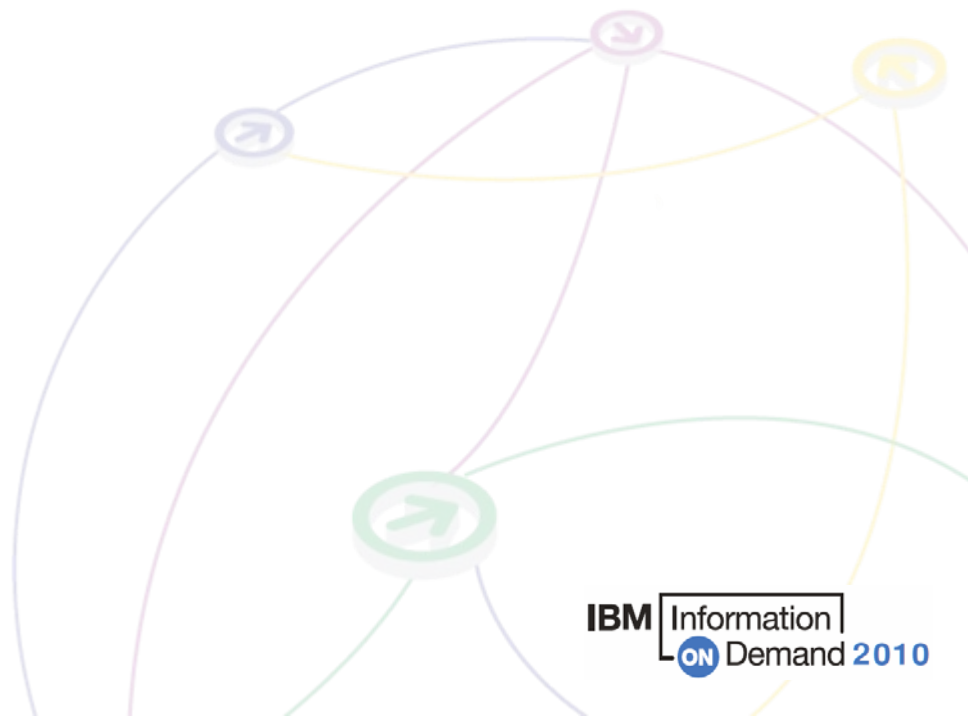
IBM Information
ON Demand 2010





Agenda

1. Vision
2. The Value of Predictive Analytics
3. Technology Follows Vision
4. Where Do We Go From Here?
5. Questions





Vision

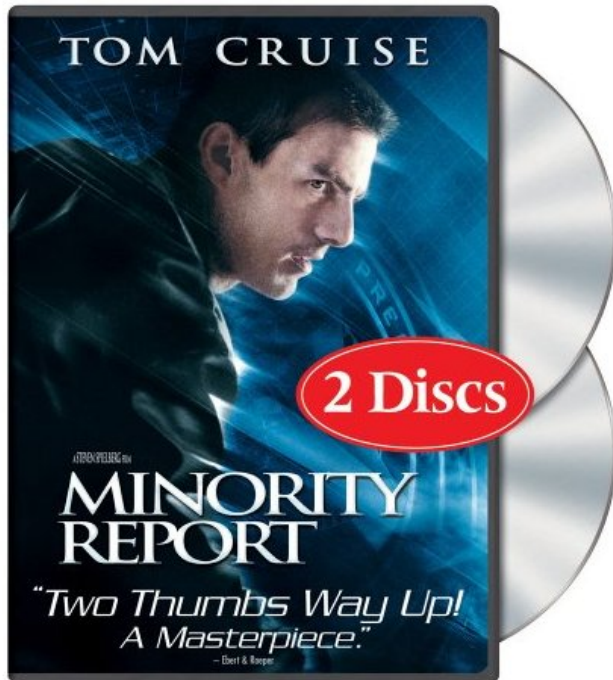
- We had a view about the transformative nature of Predictive Analytics
 - *That this...*





Vision

- *Would become this...*



Copyright © DreamWorks SKG. 2002.

Vision

→ And sure enough...

– this...



– has come to pass...



“The Future”?

→ AdsOfTheWorld.com, January 30, 2007

- “...A sign in San Francisco, starting Monday, identifies approaching Mini Cooper drivers by using a signal from a radio chip embedded in their key...the board flashes a personalized message as the driver cruises by...”

→ High Tech Billboards Tuning in to Driver’s Tastes

- San Francisco Chronicle, *“The Billboard is Listening”*
- “Roadside signs coming to Bay Area listen to car radios, then adjust pitch...”
- ...In an advertising ploy right out of Steven Spielberg's "Minority Report," electronic billboards in the Bay Area and Sacramento are being equipped to profile commuters as they whiz by -- and then **instantly personalize freeway ads based on the wealth and habits of those drivers.**
- For example, if the freeway were packed with country music listeners, the billboards might make a pitch for casinos...”

→ Google it! - There are TONS of these examples!



Creating a Predictive Enterprise

A New Source of Competitiveness

A **Predictive Enterprise**:

- Embeds analytics into key business processes
- Uses analytics to drive core business decisions
- Forward looking, not just backward!

*Life can only be understood backwards;
but it must be lived forwards.*

Soren Kierkegaard

Analytics Becomes a **Way of Life** for a Predictive Enterprise

Predictive Enterprise at Work

→ **Capture** information

- *Attributes, Interactions, Behaviors, and Attitudes*
- Customers
- Employees
- Constituents

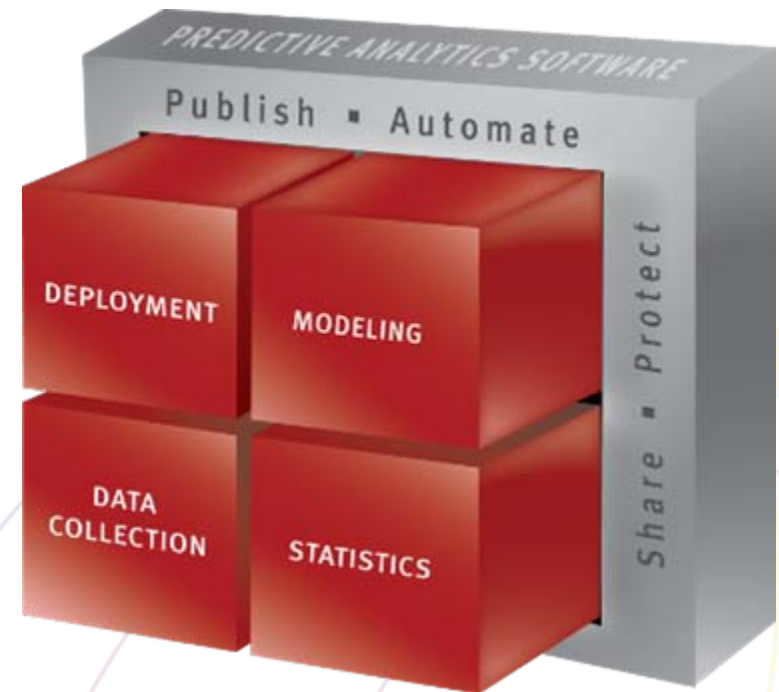
Think about
Generation Y!

→ **Predict** behavior and preferences

- Statistics for deeper insight
- Data Mining for predictive modeling
- Text Analytics for unstructured insight

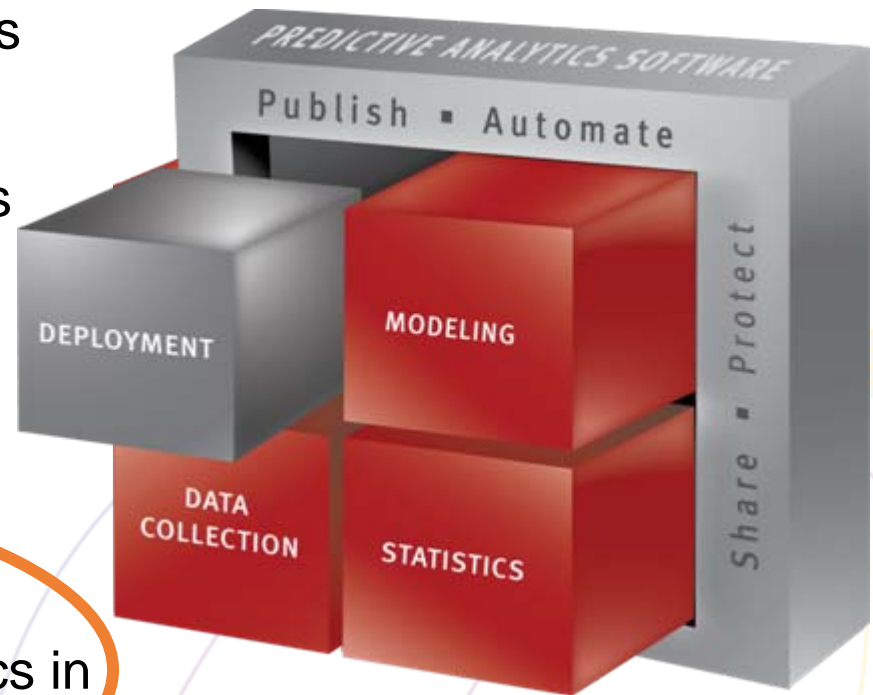
→ **Act** on results

- Efficiently deploying results
- Dramatically improving business processes



Predictive Analytics Software

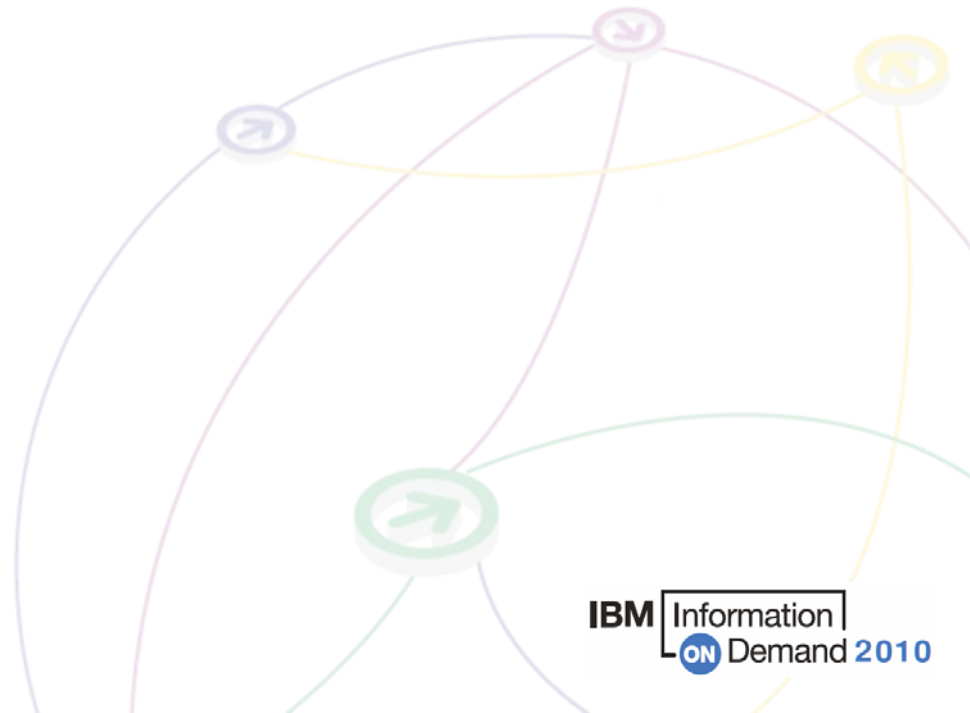
- Data Collection (capture):
 - Delivers an accurate view of customer attitudes and opinions
- Statistics (predict):
 - Drive confidence in your results and decisions
- Modeling (predict):
 - Bring repeatability to ongoing decision making
- Deployment (act):
 - Maximize the impact of analytics in your operation





Technology Follows Vision

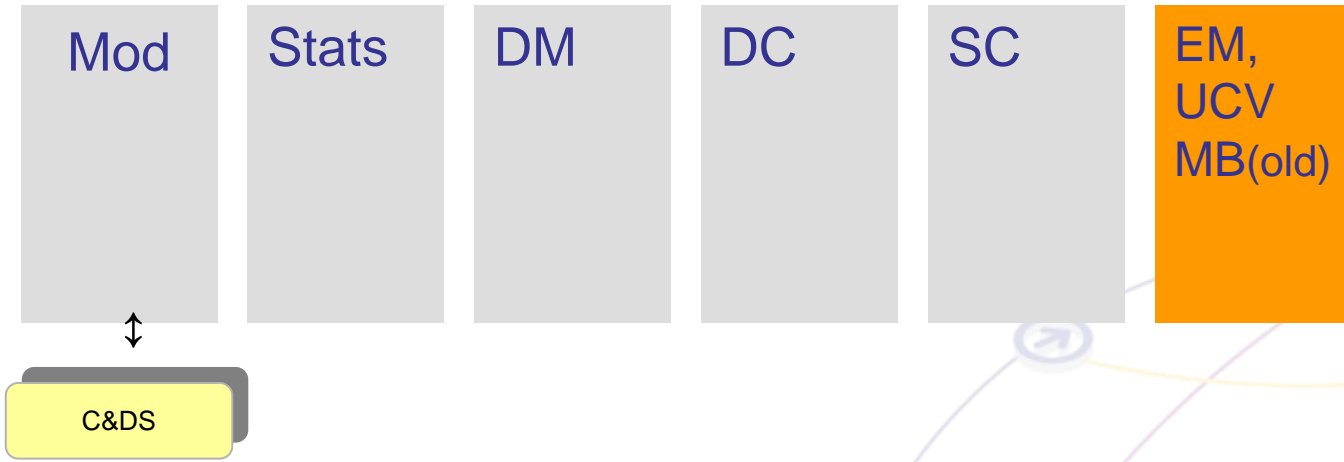
- Where we ***started***...
- Where we ***were***...
- Where we ***are***...





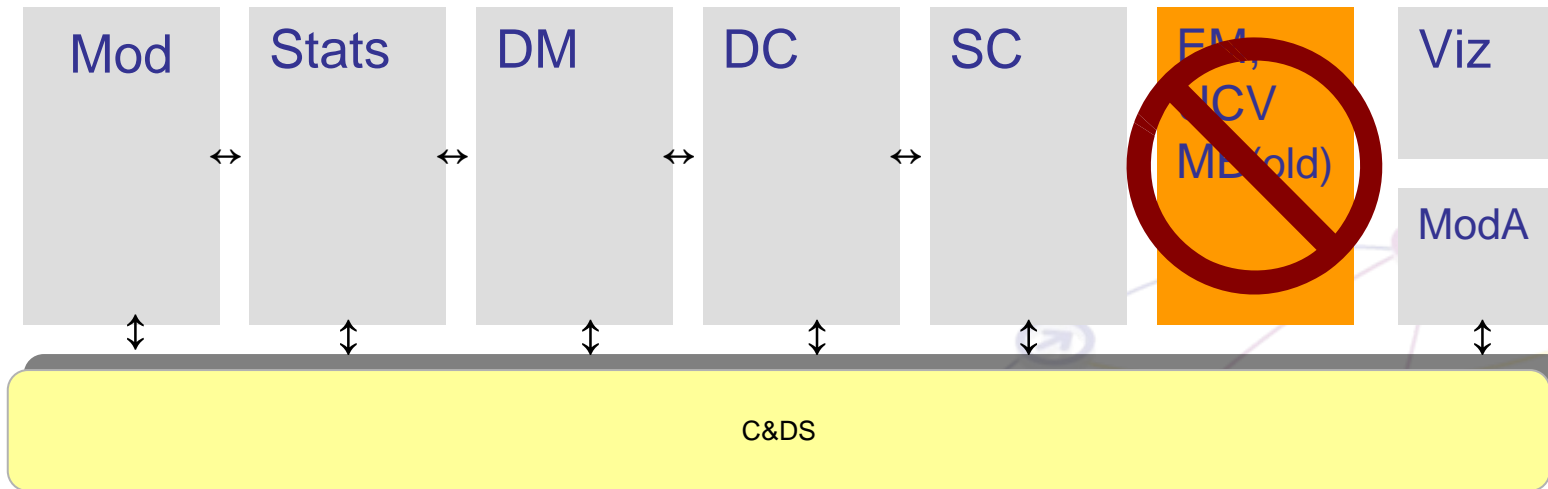
Where We Started

Silos, Closed, Trained Analysts Only



Where We Were...

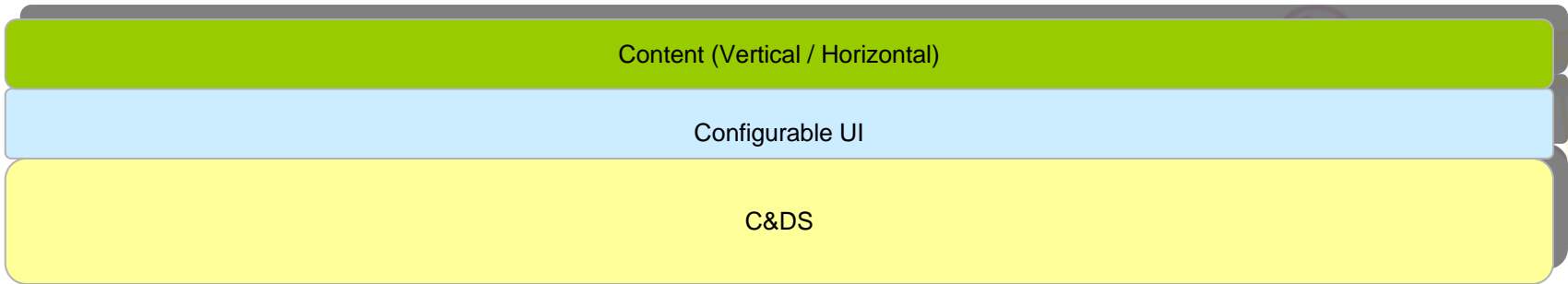
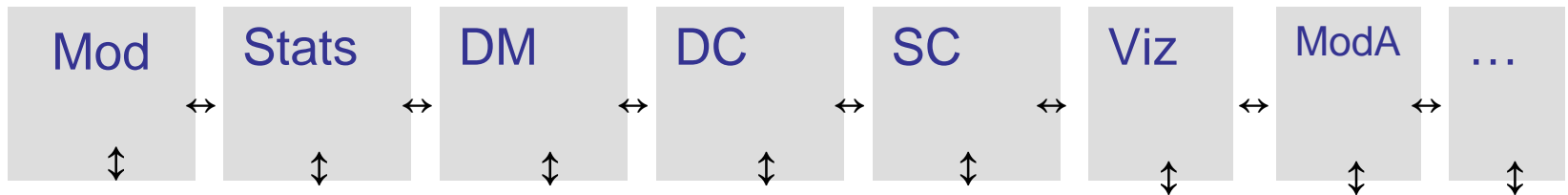
Integrated, Open, Easier, but...





Where We ARE

Easy, Content Driven Configurable

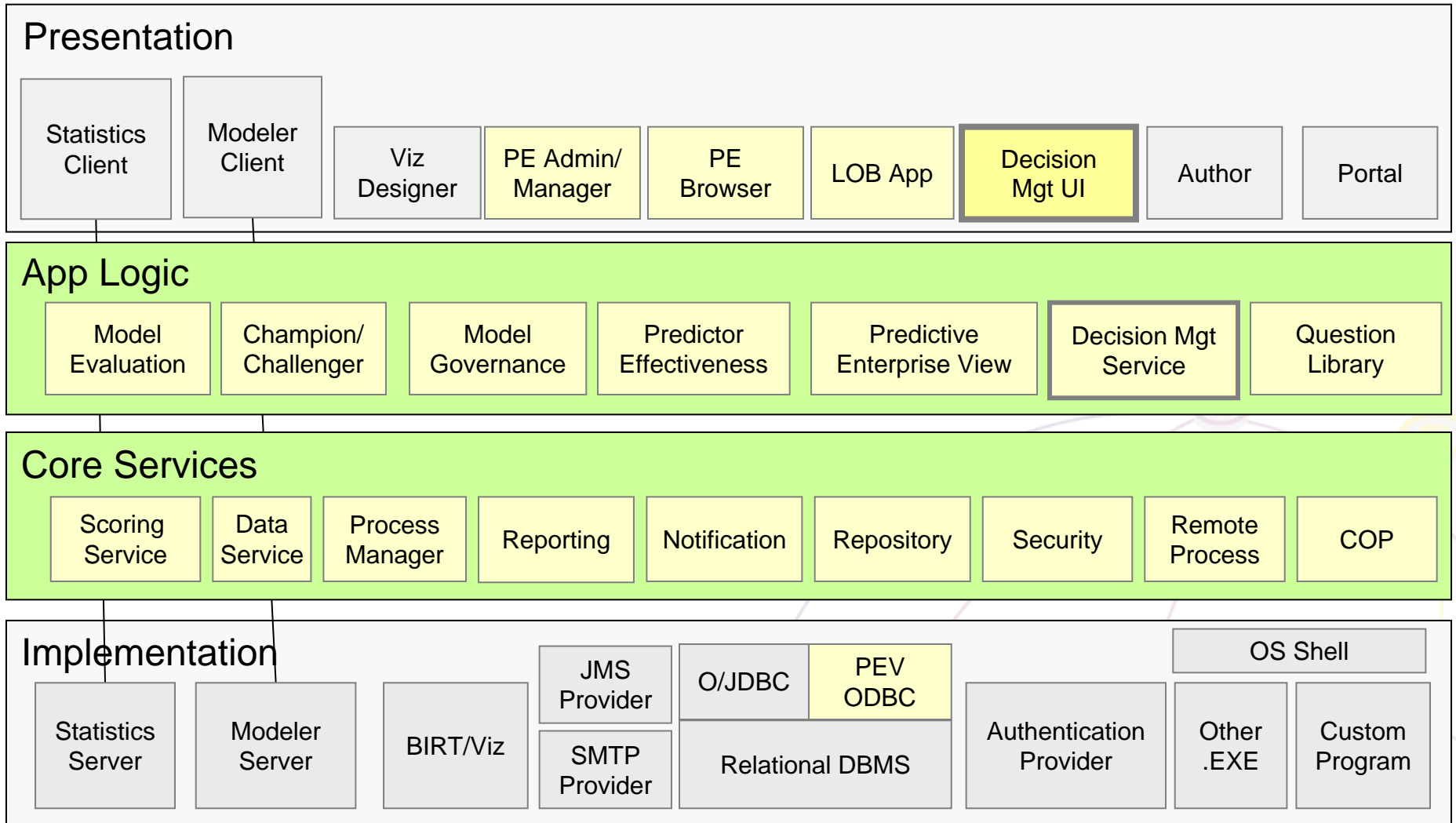




End Game We Are Driving For

- An architecture designed to create a whole new market for the mass application and deployment of Predictive Analytics at all levels
- Apply Predictive Analytics to new domains and users on the fly without 2 year R&D cycles
- Everything fits together, we leverage your environment, no investment is ever wasted, any purchase starts you on the journey

Platform Layered Architecture



Product Investment Focus

Target Users

R&D Investment Focus

- Analysts:
 - *More products for existing customers*
 - Build new add-on options
 - Build in Platform; use across multiple products
 - Cross-product integration
- “New” Analysts & Business Users:
 - *Broadening the current user base*
 - Increase usability through automation
 - Build in Platform; use across multiple products
- Consumers:
 - *Creating a new class of users*
 - Analytics are hidden from the consumer
 - Deliver deployment through the Platform
 - Cross-product integration

Science & Data

Usability

Deployment

Scalability, Integration & Openness

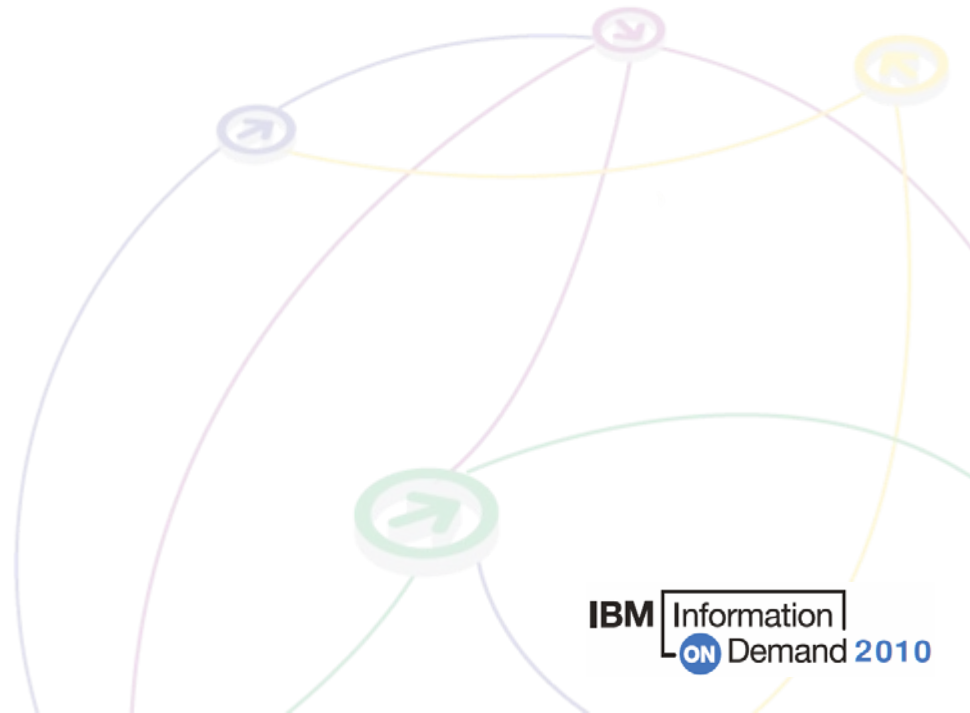
Automation & Visualization



Technology Follows Vision

→ What Capabilities are Needed to Enable the Predictive Enterprise?

- Ease of Use
- Open & Extensible
- Configurable
- Scalable
- Deployment
- Integration





Ease of Use

→ Why?

- Predictive Analytics is not just for Professional Analysts any more.
- The world will run out of highly trained analysts.

→ Product Response

- The right front end for each type of user.
- Increase the productivity of the highly trained analyst and allow others to leverage his work.

→ Example

- Building Models

Example: Building Models – Analyst vs. Business User

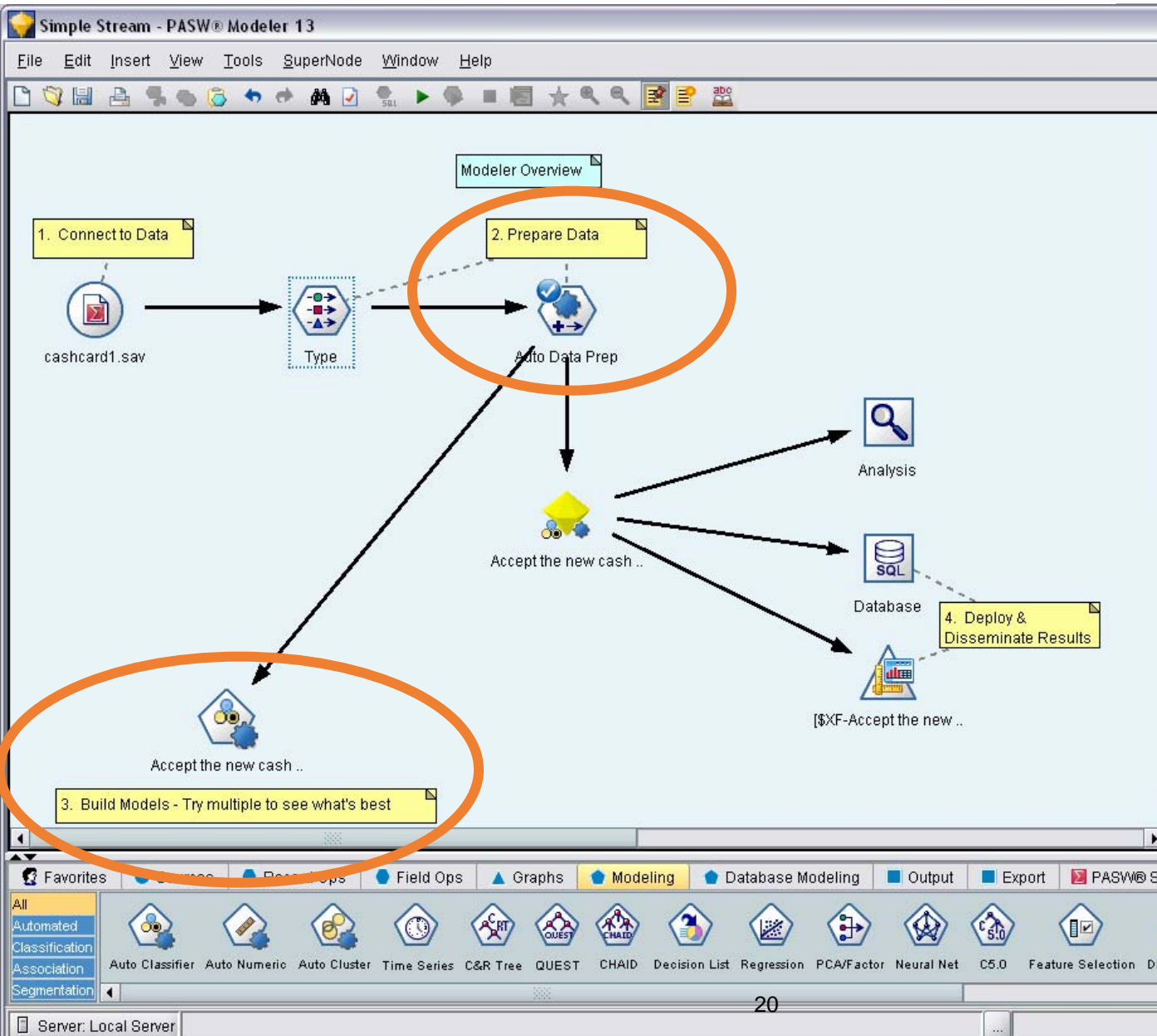
Professional Analyst:
Building a bunch of
models and
comparing results

The screenshot shows the PASW Modeler 13 interface with a workflow diagram. The workflow is divided into four main stages:

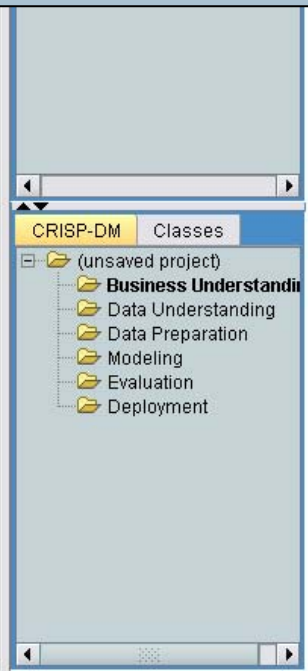
- 1. Connect to Data:** Starts with a data source 'cashcard1.sav' connected to a 'Type' node.
- 2. Prepare Data:** A sequence of nodes including 'Sort', 'Filter', '(generated)', 'SetToFlag', 'Restructure', '(generated)', and 'Filler'.
- 3. Build Models - Try multiple to see what's best:** A 'Type' node (labeled '14 Fields') branches into multiple model nodes: 'C5.0', 'CHAID', 'QUEST', and several 'Accept the new cash...' nodes.
- 4. Deploy & Disseminate Results:** An 'Ensemble' node connects to a 'Database' (SQL) and a '\$XF-Accept the new...' node.

The right-hand pane shows a project structure for 'CRISP-DM' with folders for 'Business Understanding', 'Data Understanding', 'Data Preparation', 'Modeling', 'Evaluation', and 'Deployment'. The bottom toolbar includes various modeling tools like 'Auto Data Prep', 'Type', 'Filter', 'Derive', 'Ensemble', 'Filler', 'Anonymize', 'Reclassify', 'Binning', 'RFM Analysis', 'Partition', 'SetToFlag', 'Restructure', 'Transpose', 'Time Intervals', 'History', and 'Field Reorder'.

Example: Building Models – Analyst vs. Business User



Business Analyst:
Auto Data Prepare
& Auto Model –
Easy!



on
and 2010

Example: Building Models – Analyst vs. Business User



Novice:
Press the button!

Model Type: Auto-model Interactive model

Data Source
DNB Modeling Data

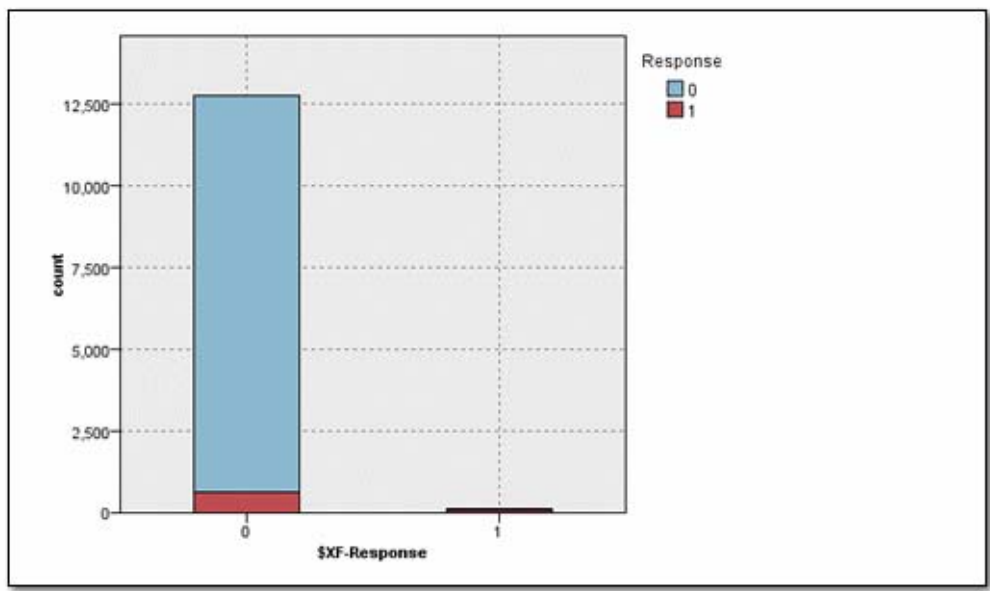
Target
Response

Build model

Evaluate

Test

Optional Settings



Distribution Chart

Distribution

Variable Importance



Examples - Open & Extensible

→ Extend Our Products

- Modeler Extension Framework
- R, Python – Add cutting edge statistical routines

→ External Control of Our Engines

- Python
- APIs

→ Standards

- Predictive Modeling Markup Language (PMML)
- In-Database Algorithms

→ External Content

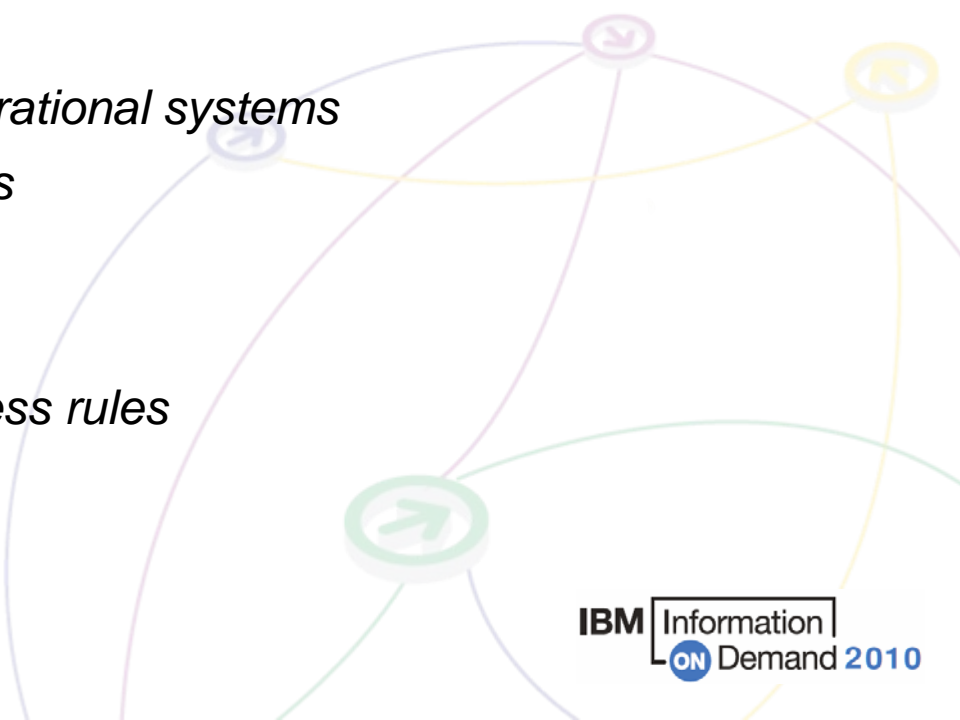
- Use third party terminology libraries and concept hierarchies to drive our Text Mining capabilities





Configurable

- Move from *Building* Solutions to *Configuring* Solutions
 - Provide across the stack content that addresses verticals and/or applications.
 - Allow that content to be further configured by Customers, Professional Services or Partners.
- For example:
 - *User interface configured to the specific application domain and users*
 - *Industry standard data models*
 - *Mappings of data to relevant operational systems*
 - *Data analysis and mining streams*
 - *Preconfigured Reports*
 - *Appropriate Visualizations*
 - *Appropriate preconfigured business rules*
 - *Pre-packaged scenarios*





Configurable

Example: PASW Decision Management

- Enables the automation of high volume, high value decision making
- Delivers recommended actions that can be deployed within an operational environments e.g. a call center or website
- Optimal decisions are delivered through the combination of predictive models, business rules, and optimization technology
- Configurable to the domain



Configurable –

Examples of Decisions Supported with Different Configurations

→ Public Security

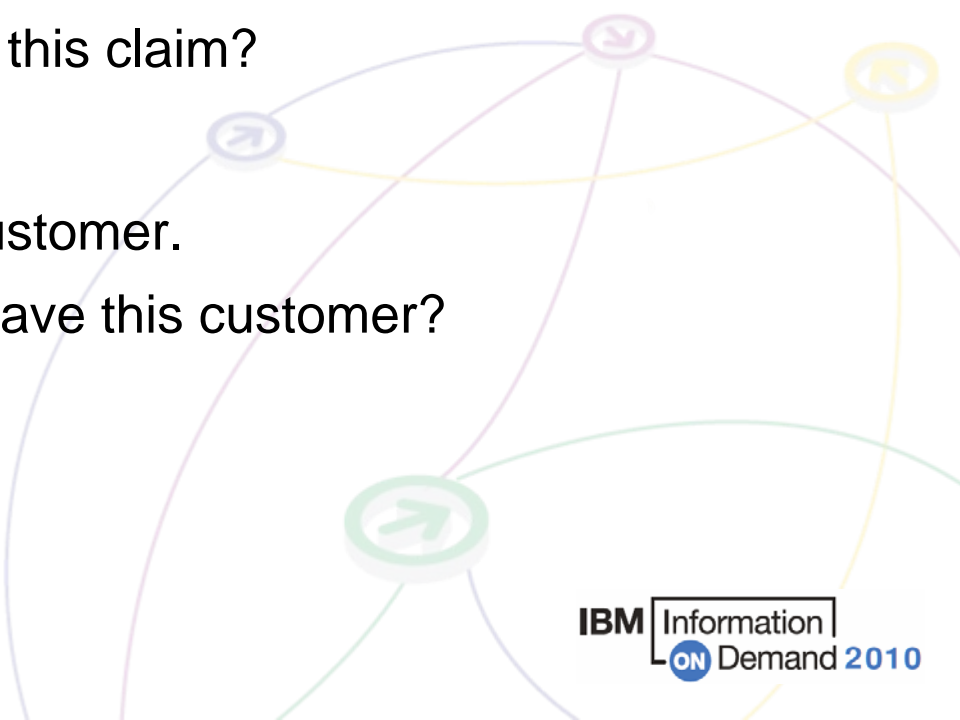
- Problem: I can't search every car that crosses the border.
- Decision: Which car should I search?

→ Insurance

- Problem: I can't investigate every claim for fraud.
- Decision: Should I investigate this claim?

→ Telecommunications

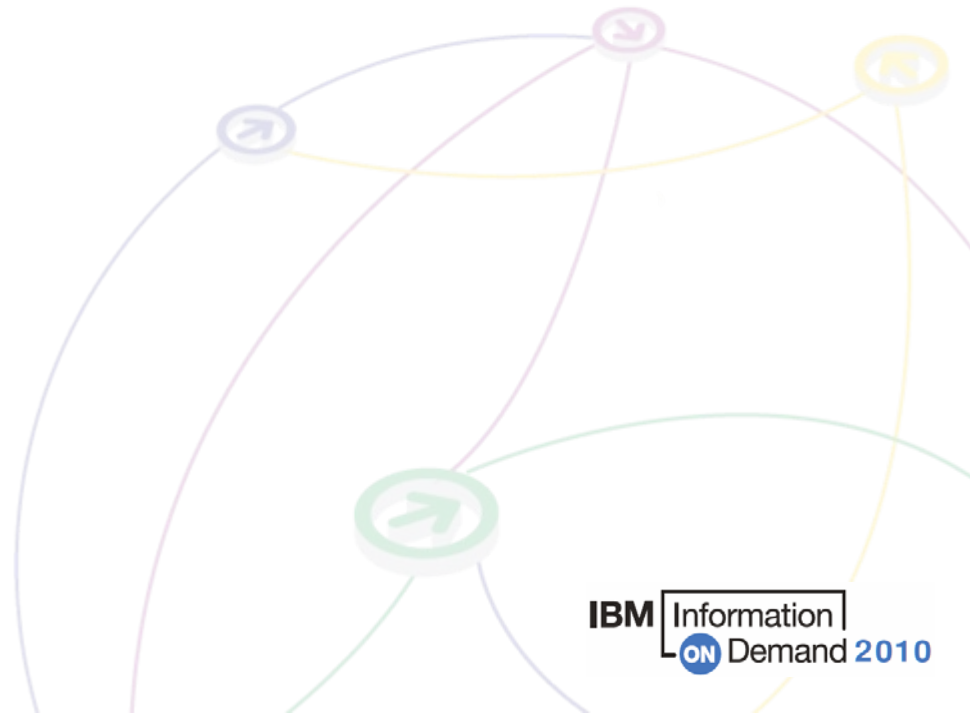
- Problem: I can't save every customer.
- Decision: Is it worth trying to save this customer?



Configurable

Example – Decision Management

1. Connect to data
2. Define enterprise-wide selections / exclusions
3. Define business rules / models
4. Combine / Optimize the decision
5. What-if / Simulation
6. Deploy / Score
7. Report



Decision Management - Call Center Configuration

→ 7 Steps – each containing configured domain specific information

Details of campaigns and offers within each campaign

Verify that the behavior is as expected on test data

High net worth customers could be identified by (1) a set of business rules or (2) a predictive model

When is the campaign available?

Which channel is it available through?

Who gets this campaign?

If the customer is valid for this campaign – which offer is presented?

Rule Name	Offer	Order
1 Female	Theatre Tickets	
2 Male	Racing Tickets	
3 Remainder	none	
4		

Scalable

→ Business Issue

- The amount of data is growing by “ridiculous” amounts.
- A large number of new different data sources is also emerging
- There is an emerging strong belief in the value of data.
- People don't like to sample!

→ SPSS Approach

- Deal with it and thrive.
- For example, In-DB Mining and PSM.

Scalable: In-Database Mining



→ Microsoft SQL Server 2005

- MS Decision Tree
- MS Clustering
- MS Association Rules
- MS Naive Bayes
- MS Linear Regression
- MS Neural Network
- MS Logistic Regression

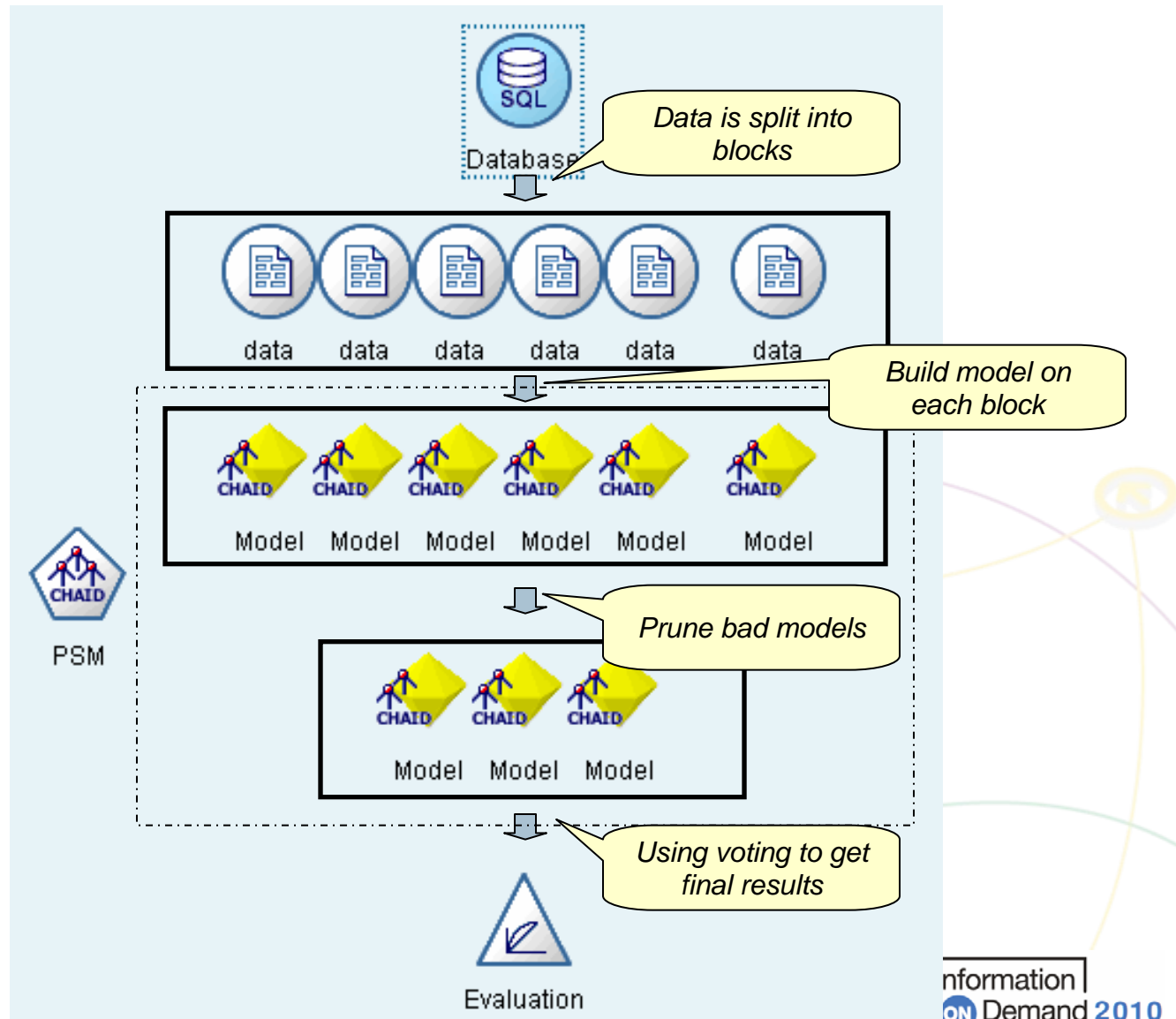
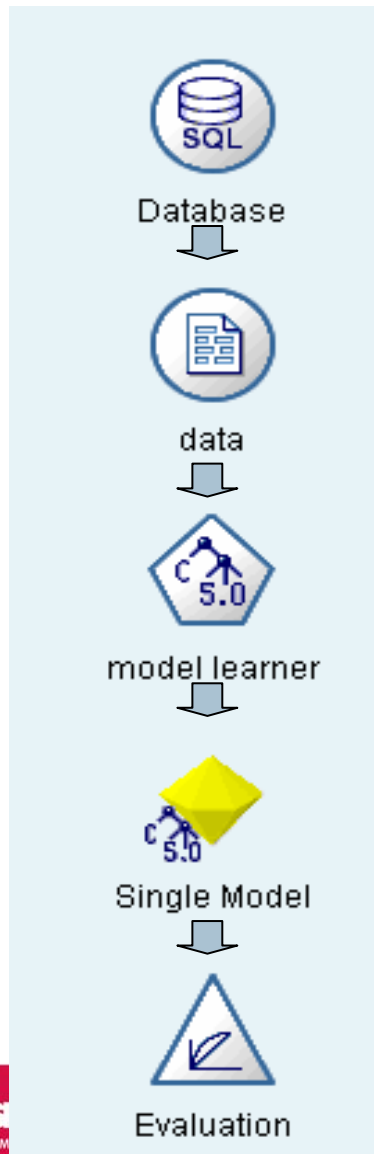
→ IBM DB2 DWE

- DB2 IM Decision Tree
- DB2 IM Association
- DB2 IM Sequence
- DB2 IM Regression
- DB2 IM Clustering

→ Oracle Data Mining

- Oracle Naive Bayes
- Oracle Adaptive Bayes
- Oracle SVM
- Oracle Decision Tree
- Oracle O-Cluster
- Oracle KMeans
- Oracle NMF
- Oracle Apriori
- Oracle MDL

Scalable: Pass Stream Merge (PSM)



Pass Stream Merge

The screenshot displays the Auto Classifier - PASW Modeler 13.0 software interface. A yellow callout box at the top left reads "An example stream with Very Large Dataset Processing". The main workspace shows a workflow starting with a file named "Credit.sav" being processed by a "Type" node, which then feeds into a "Credit risk" node. A dialog box for the "Credit risk" node is open, showing the following options:

- What is your objective?**
 - Build new model
 - Continue training existing model
- Standard Datasets**
 - Balance speed and accuracy
 - Builds a simple CHAID tree.
 - Optimize for accuracy
 - Builds an exhaustive CHAID tree.
 - Custom analysis
 - Choose this option to fine tune the algorithm on the settings tab.
- Very Large Datasets**
 - Optimize for very large datasets
 - Leverages additional processors. Requires Server connectivity.

The "Optimize for very large datasets" option is circled in red. The dialog box also includes tabs for "Objectives", "Fields", "Settings", and "Annotations", and buttons for "OK", "Execute", "Cancel", "Apply", and "Reset".

The right-hand pane shows a project tree with "Streams" (Stream1, Analyst1, RandomForest, adp, Auto Classifier, Auto Cluster, Never Finishes) and "Classes" (Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, Deployment).

The bottom toolbar includes various modeling tools such as Auto Classifier, Auto Numeric, Auto Cluster, Time Series, C&R Tree, QUEST, CHAID, Decision List, Regression, PCA/Factor, Neural Net, C5.0, Feature Selection, Discriminant, Logistic, GenLin, Cox, SVM, Bayes Net, and SLRM.

Deployment

→ *Is Deploying Predictive Analytics Scary?*





PASW Collaboration and Deployment Services (C&DS)



Collaboration

- **Create and manage analytical assets that will be used to drive the organization.**



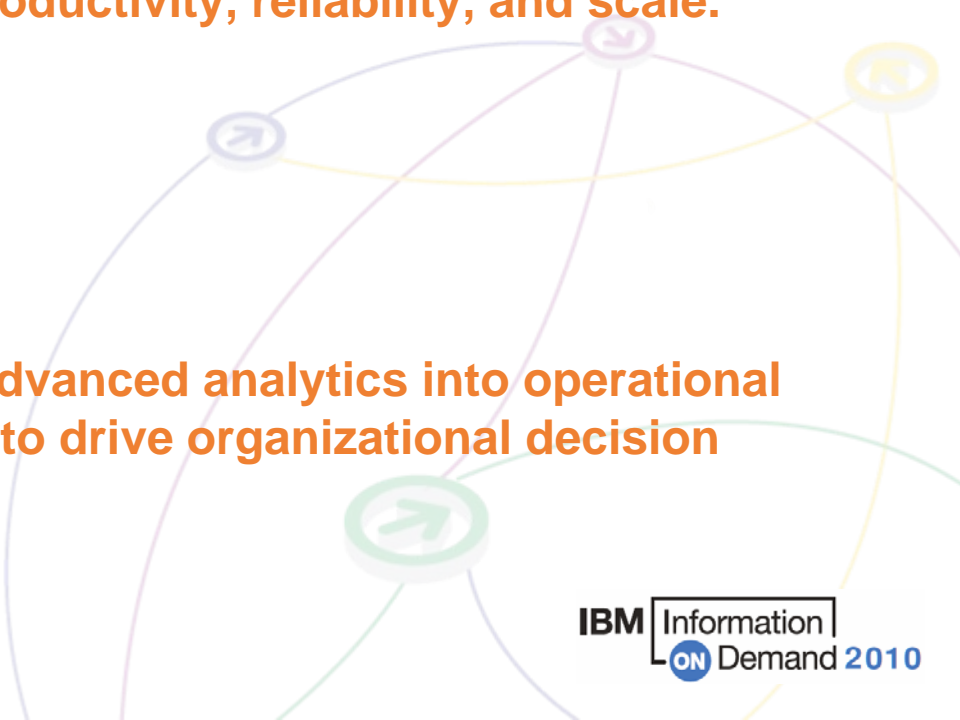
Automation

- **Increase productivity, reliability, and scale.**



Deployment

- **Deploy advanced analytics into operational systems to drive organizational decision making.**



Collaboration

→ Why is Collaboration Needed?

- Many people must be involved in the analytic process
- Control the operational cost of predictive analytics
- Compliance with regulatory and audit requirements

→ C&DS Provides

- Secure, centrally-managed asset storage
- Full lifecycle monitoring & Governance
- Productivity

Collaboration Repository of Analytical Assets

The screenshot shows a software window titled "Repository: admin@localhost". The interface includes a left-hand navigation pane with a tree view of folders such as "Boot Camp", "Case Study Files", "Exercises", "Results", "Presentations", "Samples", "Team 1-4", "Content", "Demo Data", "Directions Example", "Jobs", "Output", "Product Management", "Question Repository", "Reports", "Samples", "SAS", "Scenarios", "Scripts", "Statistics", and "Streams". The main area displays a table of files with columns for Name, Type, Size, Version, Labels, Last Modified, and Auth. A context menu is open over the file "Fraud Model.scn", showing options like "Retrieve", "Retrieve Version...", "Delete", and "Object Properties".

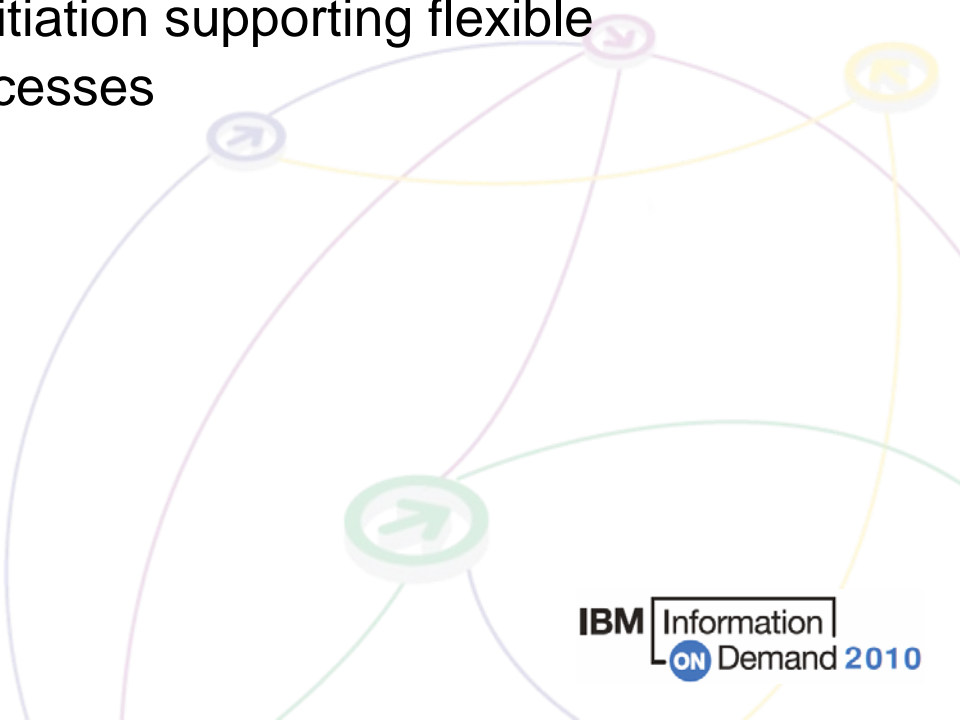
Name	Type	Size	Version	Labels	Last Modified	Auth
Champion Summary Report Object	Object	16 KB	1:2009-05-0...	(none)	Tue May 05 15:1...	user-a
Fraud Champion Evaluati...	Object	211 KB	4:2009-06-1...	(none)	Fri Jun 12 15:50...	user-a
Fraud Model.scn		32 KB	27:2009-06-...	Acceptance	Fri Jun 12 15:54...	Adminis
Fraud Model.str		33 KB	0:2009-05-0...	Test-User	Tue May 05 14:3...	Adminis
Means-User.spc		2 KB	1:2009-05-0...	(none)	Tue May 05 17:1...	user-a
Neural Prediction		25 KB	0:2009-04-2...	(none)	Tue May 05 15:5...	admin
Neural RTDPD-U		2 KB	1:2009-05-0...	(none)	Tue May 05 15:5...	user-a
NeuralScoring A			7:2009-04-2...	(none)	Tue May 05 15:4...	user-a
Peer Review Not		1 KB	1:2009-05-0...	(none)	Tue May 05 15:0...	user-a

Information for selected version:
 Name: Fraud Model.scn
 Version: 27:2009-06-12 15:54:04.718
 Labels: Acceptance
 Created By: admin
 Creation Date: Jun 12, 2009
 Keywords:
 Description: null



Automation

- Why is Automation Needed?
 - Analytical processes are complex and must be orchestrated with other systems and events
 - Manual operations are error prone and expensive
- C&DS Provides
 - Time and Event based task initiation supporting flexible orchestration of analytical processes
 - Model management



Automation: Champion / Challenger Modeling

The screenshot displays the SPSS Predictive Enterprise Manager interface. On the left, a tree view shows the project structure under '127.0.0.1 (login: admin)', including folders like 'Content Repository', 'Jobs', 'DataMining', 'ModelEvaluation', 'Reporting', and 'Survey'. The 'ModelEvaluation' folder is expanded, showing several jobs, with 'PurchaseEvaluationCCAccuracy' selected.

The central workspace shows a diagram of the modeling process. Two scenario files, 'Purchase.scn' and 'Purchase2.scn', are connected to a central box labeled 'Champion-Challenger'. Checkmarks next to the scenario files indicate they are active.

The right-hand side of the interface contains a configuration panel with several tabs: 'General', 'Challengers', 'Champion', 'Data Files', 'ODBC Data Sources', 'Parameters', and 'Notifications'. The 'Challengers' tab is currently selected, showing the following configuration:

- First Challenger:
- First Challenger Label:
- Data Provider Definition:
- Data Provider Definition Label:
- Metric:

Below the configuration fields is a table titled 'Challengers' with columns for Name, Label, and Description. The table lists four scenarios, with the first three checked:

Name	Label	Description
<input type="checkbox"/> /Scenarios/pm_demo_selflearn.scn	LATEST	
<input type="checkbox"/> /Scenarios/PM_Demo_SelfLearning.scn	LATEST	
<input checked="" type="checkbox"/> /Scenarios/Purchase2.scn	LATEST	
<input type="checkbox"/> /Scenarios/PurchaseChampion.scn	LATEST	

At the bottom right of the window, the status bar shows 'Writable' and '19M/21M'.

Deployment

→ Why is Deployment Needed?

- Analysts must have an efficient means of driving results into organizational processes
- More users require access to analytical results
- Results are required at the point of decision, sometimes in real time

→ C&DS Provides

- Real time scoring service to deliver on demand scores at the time of decision making
- Deployment of browser based configurable analysis to extend reach of analytics beyond Professional Analysts

Deployment: Scoring Integrated with Operational System

The screenshot shows a Microsoft Internet Explorer browser window titled "Scoring Sample - Microsoft Internet Explorer". The address bar displays the URL: `http://cdsdemo:8080/scoringTagLib/paswTagLib?actionName=runJSP&name=/team3/scoring.jsp&pes_userid=admin&pes_password=spss`. The page header includes the logo for "Smith Electronics Inc." and a navigation menu with buttons for "Step 1", "Step 2", "Step 3", "Step 4", and "Reports".

The main content area contains the following text:

Thanks for the interest in our financing offer.
We just need to collect a small amount of extra information to confirm if you are eligible for this offer.

The page is divided into three columns:

- PRODUCT:** LG 60inch
- STOCK & DELIVERY:** In Stock, Estimated Delivery 3 days
- Customer Details:**
 - Customer Name: Steve Johnson
 - Customer Address: 4115 Main St.
 - Customer Zip: 60699
 - Customer Email: sjohnson@web.com
 - Interest of Campaigns: Yes
- Decision:** Offer

Below the product and customer details, there is a form with the following fields:

- Age: 59
- Years at Current Address: 25
- Years at Current Employer: 21
- Education Level: College
- Income (in thousands): 180
- Credit Card Debt (in thousands): 20
- Other Debt (in thousands): 0
- Previous Loan Default: No

A "Run Score" button is located at the bottom left of the form area. A red circle highlights the "Decision" box, and a yellow circle highlights the "Run Score" button.

Predictive Analytics - A Journey

Bullseye!

Fundamental organizational transformation achieved... yielding true competitive advantage.



Claims Processing

Data Global Selections Define Combine Deploy Reports

Motor & Home Testing

Motor Claims

Allocated actions for Motor Claims will override matrix results

Use same matrix for all interaction points All interaction points

Rules Action	Combine Matrix	Model Actions		
		High	Medium	Low
Refer to SIU	Refer to SIU	Standard Processing	Standard Processing	Standard Processing
Standard Processing	Refer to SIU	Standard Processing	Standard Processing	Fast Track
Fast Track	Standard Processing	Standard Processing	Standard Processing	Fast Track

Matrix Colors

- Refer to SIU
- Standard Processing
- Fast Track
- High
- Medium
- Low

What if ...?

Test...

Mark as done

The Predictive Enterprise: Example

Cable Company Customer Retention

→ Background

- Leading Swiss provider of telephony, internet and cable TV services

→ Business goals

- Reduce churn among broadband and internet customer base
- Identify customers likely to cancel in time to proactively address concerns / dissatisfaction

→ SPSS Portfolio drives success

- Use PASW Statistics for valued customer analysis across 60 key variables
- Use PASW Modeler on transaction data to predict churn
- Add text data and mining to better predict customer churn and understand reasons
- Use PASW Data Collection for customer satisfaction measurement.
- Customer Lifetime Value metric integrated into call center applications

SPSS Technology

Data Analysis & Workbenches

- Modeler
- Statistics

Attitudinal Data

- Data Collection
- Text Mining

Collaboration & Deployment Services

- Analytic insight mgt
- Automated analysis



**“Reduced churn
from 19 to 2%”**

Summary: Are you ready for tomorrow?

→ **Yes !!! Tomorrow is here!**

→ Non-invasive.

→ Supports the Journey.

→ Anyone can do it.

→ Configurable.

→ Manageable.

→ Adds value that transforms an organization.



Questions



Alex Lee

Technical Sales Lead/IT
Specialist

leeyma@sg.ibm.com

Please visit the SPSS booth for quick demo or more info.
Thank you for your time.



Thank You!

Your Feedback is Important to Us

- Please complete the survey for this session by:
- Accessing the SmartSite on your smart phone or computer at:
iodsmartsite.com
 - Surveys / My Session Evaluations
 - Visiting any onsite event kiosk
 - Surveys / My Session Evaluations
 - Each completed survey increases your chance to win an Apple iPod Touch with daily drawing sponsored by Alliance Tech

