IBM Persistent Analytics Demo

Bernie Beekman
Analytics Solution Center
Agenda

- Demo Objectives
- Demo Components
- Demo Use Cases
- Demo
- Future Work
DEMO OBJECTIVES
Demo Objectives

- **Implement a Fraud and Improper Payments Prevention use case**
  - Addresses government pain point
  - Fraud impacts multiple agencies.

- **Illustrate the seamless integration of IBM products in an end-to-end fashion**

- **Ensure the demo will be flexible, and readily adapted to reflect the mission needs of specific agencies across Federal**

- **Ellicit feedback from stakeholders via hosting at the Analytics Solution Center (ASC).**
Initial Demo Capabilities

- Real Time Fraud Detection
- Identity and Relationship Resolution
- Predictive Analytics Engine
- Data Pipeline
- Enterprise Dashboard
Demo System Context

Entities → Analyze

Transform → Store

Visualize

Transactions

Analyst User

Executive User

User
Demo Components

- Entities
- Transactions
- Data Integration
  - InfoSphere Information Server
- Analytics Engine
- Data Repository
  - DB2
- Visualization
  - Cognos Business Intelligence
  - Identity Insight Visualizer
- Users
  - SPSS Modeler
  - Identity Insight
Demo Components

- Cognos Business Intelligence
- SPSS Modeler
- Identity Insight
- InfoSphere Information Server
- DB2
Cognos Business Intelligence

- Unified Workspace
- Dashboards
- Reports
- Scorecards
- Real-Time Monitoring
- Event Alerting
- Analysis and Exploration
- Ad-hoc Query
- Trend and Statistical Analysis
- Predictive Analysis
- Role-Based
- Framework Manager
Demo Components

- Cognos BI
- SPSS Modeler
- Identity Insight
- InfoSphere Information Server
- DB2
Leveraging Information for Smarter Organizational Outcomes

SPSS Modeler – Predictive Analytics
Demo Components

- Cognos BI
- SPSS Modeler
- Identity Insight
- InfoSphere Information Server
- DB2
IBM InfoSphere Identity Insight Discovers “Who is Who”
IBM InfoSphere Identity Insight Discovers “Who knows Who”

Relationship Resolution creates the social network which can then be used for real-time threat assessment and investigatory link analysis.
Alerts Tell You When Something is Wrong

Relevance finds the user

Role Alert: When entities are resolved or related

Examples
- Employee is a vendor
- Employee shares address with vendor
- POI shares last name and phone number with police officer
- Employee shares bank account with employee
- Person has more than one name

Attribute Alert: When an entity with specific set of attributes is found

Examples
- If a name is ever encountered
- When an address of a suspect is ever encountered
- When any collection of one or more attributes is found

Event Alert: When transactions and identities match alert rules

Examples
- Five money transfers in 24 hours over $10,000 by one person
- More than one account opened in 2 days by same person
- 10 or more code violations reported in 5 days by same inspector
InfoSphere Identity Insight
Unique Analytical Abilities to Highlight Suspicious Situations

- **Full Attribution**: A full history of each individual is accumulated retaining all attributes.
- **NORA™**: Non-Obvious Relationship Awareness allows for conflict detection at 2 to N degrees of relationship separation.
- **Real-time Persistent Search**: Alerts sent the moment that data an investigator has registered about a person or set of attributes is introduced to the system.
- **Multi-Level Event Monitoring**: Monitor for suspicious events as they relate to individual sources, resolved entities and relationships.
- **Perpetual Analytics**: Un-resolve/Re-resolve, sequence neutral. It works in real-time on an ongoing basis. Identity information is perpetually updated.
- **Multi-cultural Name Recognition**: Automatic application of linguistic rules that finds matches with regard to specific cultural context via patented IBM technology.
- **Insight/Alerting at Ingestion**: Performs all analytics at ingestion on data streams for real-time discovery and alerting.
- **Role Conflicts**: User configured rules to provide alert notification whenever data connects conflicting roles.
Demo Components

- Cognos BI
- SPSS Modeler
- Identity Insight
- InfoSphere Information Server
- DB2
What is the Information Server?
InfoSphere DataStage

Integrate, transform and deliver data on demand across multiple sources and targets including databases and enterprise applications.

Requirements
- Integrate and transform multiple, complex, and disparate sources of information.
- Demand for data is diverse – DW, MDM, Analytics, Applications, and real time.

Benefits
- Transform and aggregate any volume of information.
- Deliver data in batch or real time through visually designed logic.
- Hundreds of built-in transformation functions.
- Metadata-driven productivity, enabling collaboration.
Work Flow

Integration Landscape

Plan/Design

Deploy & Optimize

Manage

InfoSphere™
Foundation Tools

Cleanse Data

Transform Data

Optimized

TARGETS

BI

Warehouse

MDM

ODS

Etc...
DEMO USE CASES
Demo Use Cases – Tax Fraud

1. **Bad Tax Preparer** - A 1040 Form is submitted with a seemingly legitimate Tax Preparer. However, the Identity Insight Engine determines a relationship exists between the Tax Preparer and a known Bad Tax Preparer causing a role alert.

2. **Head of Household** - A 1040 Form is submitted with a Head of Household who is a dependent on another tax form. The business rule that detects this is implemented in SPSS Modeler.

3. **Anomalous Behavior** - Historical Tax Returns are analyzed to determine what would be categorized as an anomaly. The model that results from this analysis is *deployed* to score incoming transactions to determine if they are anomalous.

4. **Executive Dashboard** – A Dashboard is created to summarize fraud statistics geographically and by department.
Bad Tax Preparer Use Case

- Based on an actual DOJ Case
- Tax preparer was barred from doing business due to the filing of fraudulent tax returns.
- Reported 3 Businesses that he owned.
- Identity Insight discovered 6 additional as part of the Tax Preparer registration process.
- Identity Insight flags returns filed by all 9 companies for further review due to affiliation with a known fraudulent tax preparer.

Reported these businesses to DOJ

II also discovered relationships with 6 additional businesses

Victor Bondarenko

Tax Express
Tax Express Inc.
Tax Cash Express

EC Tax and Business Services
Fast cash Refund Express
Electronic Tax Refund Express

......
Leveraging Information for Smarter Organizational Outcomes

Tax Fraud Analytics Demo Architecture

Step 1: Parse, Cleanse and Transform Data
- InfoSphere Server
  - DataStage
  - QualityStage
  - Analyzer
  - Metadata Server

Pre-condition: Entity Data has been loaded:
- Bad Tax Preparer List
- Tax Preparation Firms
- ...

Identity Insight

Step 2: Resolve Identities and Relationships

Step 3: Store Transaction Data in the Warehouse

Step 4: Perform Predictive Analytics
- SPSS Modeler Server
  - DB2 Alerts

Cognos
- Reports
- Executive Dashboard

Demo Driver Servlet
Reset DBs
Launch demo

© 2009 IBM Corporation
FUTURE EXTENSIONS
Future Extensions

- Entities
- Data Integration (InfoSphere Information Server, InfoSphere Streams)
- Analytics Engine
- Data Repository (DB2, Netezza)
- Visualization
- Users (SPSS Modeler, Identity Insight, WODM, GNR, Cognos Business Intelligence, I2, IBM Case Manager)
Future Extensions

- **I2** - provides advanced visualization to support investigations

- **Advanced Case Management** – provides case file creation, workflow management, and analytics for investigations

- **InfoSphere Streams** – scalable engine analyzing high volume input and providing real time alerting. Suitable for Threat Prediction and Prevention use case.

- **WODM** – provides a rules development, run time, and management capability

- **Global Name Recognition** - helps manage, search, analyze, and compare multicultural name data sets by leveraging culture-specific name data & linguistic rules.

- **Netezza** – data warehouse appliance purpose-built for analytics.

- More Use Cases …
Future Extensions: Anomalous Behavior

Use SPSS Advanced Analytic Capabilities to Detect Potentially Fraudulent Returns

- Anomalous Data Values Across Multiple Attributes
- Cluster Analysis
- Historical Precedence Analysis
- Business Rule Set Analysis
- ...
Organizations need answers to key questions.

- Operations
- Product Development
- IT
- Finance
- Marketing
- Sales
- Customer Service
- HR
How are we doing?

**Immediate Insights to Business Performance**

- Dashboards
- Reports
- Scorecards
- Real-Time Monitoring
- Event alerting
Leveraging Information for Smarter Organizational Outcomes

Why?

Deeper Analysis of Trends & Patterns

- Analysis and Exploration
- Ad-hoc Query
- Trend and Statistical Analysis
What should we be doing?

Foresight to Plan & Allocate Resources

- What-If Analysis
- Predictive Analysis
Transforming Business Intelligence

Broad and Powerful Analytics

- Scorecards
- Dashboards
- Ad-hoc Query
- Real-Time Monitoring
- Reports
- What-If Analysis
- Trend & Statistical Analysis
- Predictive Analysis
- Analysis & Exploration

Unified Workspace

Proven Platform

Cognos BI
Identity Resolution

Data types used to uncover “Who is Who”

Name
- Last name
- First name
- Middle name
- Other name parts
- Generation
- Org name
- Aliases

Identifiers
- Credit card
- Driver’s license
- Bank account
- Tax ID
- Passport
- Loyalty club
- Phone

Attributes
- Date of birth
- Circa date of birth
- Nationality
- Citizenship
- Place of birth
- Height, Weight
- Eye, Hair color

Location
- Address
- City
- State/province
- Postal code
- Country
- Latitude/longitude

Digital
- Email address
- Cookie
- IP address
- Audio
- Video

Events
- Criminal Activity
- Phone Calls
- Money Transfers
- Crossing Border

Affiliations
- Organized Crime
- Gang
- Terrorist
- Sympathetic Organizations