Unleash the Value of Big Data through Predictive Analytics with SAP HANA

Philip Mugglestone
SAP
Learning Points

- New market forces are changing the landscape and offering new opportunities
- Predictive analysis can be for everyone in the business
- SAP HANA is the cornerstone of SAP’s vision for predictive analytics
Agenda

- The Predictive Analytics Landscape
- Predictive Analytics with SAP HANA
- Customers Benefit from Predictive Analytics with SAP HANA
- Predictive Applications for SAP HANA
- Demo
Agenda

- The Predictive Analytics Landscape
- Predictive Analytics with SAP HANA
- Customers Benefit from Predictive Analytics with SAP HANA
- Predictive Applications for SAP HANA
- Demo
What if…

. . . You could identify hidden revenue opportunities within your customer base through predictive analytics?

. . . You could retain your high-value customers/employees/vendors/partners with the right retention offers?

. . . Your call center agents could delight customers with the best next-step recommendations?

. . . You could increase cross-sell and up-sell effectiveness through cross-channel coordination?

. . . You could build long-term customer/employee/vendor/partner relationships with intelligent interactions?
Extend Your Analytics Capabilities

The key is unlocking data to move decision making from sense & respond to predict & act
What is Predictive Analysis?

Predictive analysis encompasses a range of analytic techniques

“... the exploration and analysis, by automatic or semi-automatic means, of large quantities of data in order to discover meaningful patterns and rules.”

Gordon Linoff and Michael Berry
Authors of “Data Mining Techniques”

“... the process of discovering meaningful new correlations, patterns and trends by sifting through large amounts of data stored in repositories, using pattern recognition technologies as well as statistical and mathematical techniques.”

Gartner Group
Predictive Analytics
Examples

**Challenges**

- **Forecasting**
  - How do historical sales, costs, key performance metrics, and so on, translate to future performance? How do predicted results compare with goals?

- **Anomalies**
  - What anomalies might exist and conversely what groupings or clusters might exist for specific analysis?

- **Key Influencers**
  - What are the main influencers of customer satisfaction, customer churn, employee turnover, and so on, that impact success?

- **Relationships**
  - What are the correlations in the data? What are the cross-sell and up-sell opportunities?

- **Trends**
  - What are the trends: historical / emerging, sudden step changes, unusual numeric values that impact the business?
Changing Landscapes and New Opportunities

Data mining and predictive have been around for decades. But new market forces are changing the landscape and offering new opportunities...

- **Increased business interest**
  - Now that BI users know what happened, they are asking why and what’s likely to happen next
  - Explosive demand from sales, marketing, and call center analyses... fraud, and government intelligence/security agencies

- **Increased data value (e.g., Big Data)**
  - Exploding data volume
  - Expanding data varieties

- **Increasing technology performance**
  - Parallel processing, faster CPUs, and in-memory technologies reduce time and cost of data processing
Big data matters
Transformational business value from data

Business Value
- Drive Better Profit Margins
- Operational Efficiencies
- New Strategies and Business Models

Velocity
- CRM Data
- Customer
- Things
- Sales Order

Volume
- Instant Messages
- Mobile
- Demand
- Transactions

Variety
- Planning
- Inventory
- Opportunities
SAP Predictive Analytics Vision

Database to Decision Predictive Analytics
- In-database predictive and R integration
- Intuitive predictive modeling
- Advanced data visualization and easy to use exploration

For Everyone in the Business
- Seamlessly embedded in business user applications
- Extended into BI clients and reports
- Insight into events instantly delivered to dashboards, alerts, and mobile devices

Harnessing Powerful Big Data Analytics
- Real-time on massive amounts of structured and unstructured data
- Complex questions answered in-memory, lightning fast
- Deep Hadoop integration with built-in text analysis
Predictive Analysis - User Personas

Number of users

- 5 / 20
- 50 / 100
- 500 / 5,000

User personas

- Application End User
  - Information Consumer
  - Interactive Consumer
- Business Analyst / Interactive Consumer
- Business Analyst
- Professional Data Analyst
- Predictive Analysis process designer / author

Embedding Predictive Analysis
- Industry applications
  - LOB applications
  - BI client tools
- All personas
- Large number

SAP HANA
- PAL & R
- IQ
- SAP PA designer
- SAP PA wizard
- SAP PA visualization

© 2012 SAP AG. All rights reserved.
Agenda

- The Predictive Analytics Landscape
- Predictive Analytics with SAP HANA
- Customers Benefit from Predictive Analytics with SAP HANA
- Predictive Applications for SAP HANA
- Demo
Predictive Analytics with SAP HANA
Transforming the Future with Insight Today

Unleash the value of Big Data through the power of SAP HANA
- Employ in-database predictive algorithms
- Access 3,500+ open-source algorithms via R integration for SAP HANA

Intuitively design and visualize complex predictive models
- SAP Predictive Analysis software

Bring predictive insight to everyone in the business
- Embed within business applications
- Extend into BI and reports
- Insight into events instantly delivered to dashboards, alerts, and mobile devices
SAP HANA In-Memory Predictive Analytics

Combine the depth and power of in-memory analytics within SAP HANA with the breadth of R to support a variety of advanced analytic and predictive scenarios.

Predictive Analysis Library (PAL)
- **Native** predictive algorithms
- In-database processing for **powerful** and **fast** results
- **Quicker** implementations
- Support for K-Means, K-Nearest Neighbor, C4.5 Decision Tree, Multiple Linear Regression, Apriori, ABC Classification, Weighted Score Tables

R Integration for SAP HANA
- Enables the use of the **R open source environment** (> 3,500 packages) in the context of the HANA in-memory database
- R integration enabled via **high performing** parallelized connection
- R script is **embedded** within SAP HANA SQL Script
# SAP HANA In-Memory Predictive Analytics

## Predictive Analysis Library (PAL)

Additional native predictive analysis functions provide for deeper insights and faster analytic implementations as well as results.

### SAP HANA SPS3

- **Clustering**
  - K-Means

- **Classification**
  - K-Nearest Neighbor
  - C4.5 Decision Tree
  - Multiple Linear Regression

- **Association**
  - Apriori

- **Classification**
  - ABC Classification
  - Weighted Score Tables

### SAP HANA SPS4

- **Association**
  - Lite Apriori (A -> B)

- **Classification**
  - Exponential / Geometric / Logistic / Natural Logarithmic Regression
  - CHAID Analysis

- **Time Series**
  - Single / Double / Triple exponential smoothing

- **Preprocessing**
  - Outlier Detection (Inter-Quartile Range)

- **Project Solution**
  - See SAP Note 1705650

- **The Predictive Analysis Library (PAL) is part of the Business Function Library**
- **It resides in the Calculation Engine and consists of functions executing at the database layer and is written in C++**
R Integration for SAP HANA
What is R?

R is a software environment for statistical computing and graphics
- Open Source statistical programming language
- Over 3,500 add-on packages; ability to write your own functions
- Widely used for a variety of statistical methods
- More algorithms and packages than SAS + SPSS + Statistica

Who’s using it?
- Growing number of data analysts in industry, government, consulting, and academia
- Cross-industry use: high-tech, retail, manufacturing, CPG, financial services, banking, telecom, etc.

Why do they use it?
- Free, comprehensive, and many learn it at college/university
- Offers rich library of statistical and graphical packages
R Integration for SAP HANA
Adoption – R is used by the most data miners

- Data miners report using an average of 4.6 software tools.
- R is used by the most data miners (43%).
- STATISTICA is the primary data mining tool chosen most often (18%).
**R Integration for SAP HANA**

*Functionality Overview*

**R integration for SAP HANA** enables the use of the R open source environment in the context of the HANA in-memory database

- Establishes a communication channel between HANA and R for fast data exchange
- Improved data exchange mechanism supports transfer of intermediate database tables directly into vector oriented data structures of R
- Performance advantage over standard tuple-based SQL interfaces with no need for data duplication on the R server
- Supported by providing the R-operator as a custom operator in the calcModel of the calculation engine
- This allows the application user to embed R script within SQL script and submit entire query to the HANA database
- As the plan execution reaches R-node, a separate R runtime is invoked using Rserve and input tables of R node passed to R process using improved data transfer mechanism
Predictive Analytics Process in SAP HANA

Step 1: Data Loading
1. Understand the business and figure out the problems
2. Load the SAP or non-SAP data into SAP HANA

Step 2: Data Pre-Processing
1. Data selection
2. Data cleansing
3. Data transformation

Step 3: Data Mining
1. Train Predictive Model (clustering / classification / association / time series etc.)
2. Predict based on suitable model

Step 4: Visualization
1. Visualize the model for better understanding
2. Store the model and results in SAP HANA
Agenda

- The Predictive Analytics Landscape
- Predictive Analytics with SAP HANA
- Customers Benefit from Predictive Analytics with SAP HANA
- Predictive Applications for SAP HANA
- Demo
**Business Challenges**
- Reduce delays and minimize the costs associated with new drug discovery by optimizing the process for genome analysis
- Improve and speed decision making for hospitals which conduct cancer detection based on DNA sequence matching

**Technical Implementation**
- Leveraged the combination of SAP HANA, R, and Hadoop to store, pre-process, compute, and analyze huge amounts of data
- Provide access to breadth of predictive analytics libraries

**Benefits**
- For pharmaceutical companies, provide required new drugs on time and aid identification of “driver mutation” for new drug targets
- Able to provide a one stop service including genomic data analysis of cancer patients to support personalized patient therapeutics

---

Our solution is to incorporate SAP HANA along with Hadoop and R to create a single real-time big data platform. With this we have found a way to shorten the genome analysis time from several days down to only 20 minutes.

*Yukihisa Kato, CTO and Director of MITSUI KNOWLEDGE INDUSTRY*
The HANA platform at Cisco has been used to deliver near real-time insights to our execs, and the integration with R will allow us to combine the predictive algorithms in R with this near-real-time data from HANA. The net impact is that we will be able to take the capability which takes weeks and months to put together, and deliver just-in-time as the business is changing.

Piyush Bhargava, Distinguished Engineer IT, Cisco Systems
Bigpoint
Gaming Industry - Predictive Game Player Behavior Analysis

Business Challenges
• Increase conversion rates from free → paying player
• Increase the average revenue per paying player
• Decrease churn – keep paying players playing longer

Technical Challenges
• Leverage real-time data processing in SAP HANA and classification algorithms with R integration for SAP HANA to deliver personalized context-relevant offers to players
• Analyze vast amounts of historical and transactional data to forecast player behavior patterns

Benefits
• Real-time insights
• Per player profitability analysis and increased understanding of player behavior
• Increase data volume and processing capabilities to communicate personalized messages to players

At Bigpoint in the Battlestar Galactica online game, we have more than 5,000 events in the game per second which we have to load in SAP HANA environment and to work on it to create an individualized game environment to create offers for them. In this co-innovation project with SAP HANA, using Real Time Offer Management Bigpoint, we hope to increase revenue by 10-30%.

Claus Wagner, Senior Vice President SAP Technology, Bigpoint
Agenda

- The Predictive Analytics Landscape
- Predictive Analytics with SAP HANA
- Customers Benefit from Predictive Analytics with SAP HANA
- Predictive Applications for SAP HANA
- Demo
SAP HANA Predictive Ecosystem

SAP HANA Platform

Data Pre-Processing and Loading
SAP Data Services, Information Composer, SLT, DXC, Hadoop
SAP HANA – Hadoop Support in SAP Data Services 4.1
Extend the “Big Data” Ecosystem to SAP HANA

- SAP HANA can integrate with Hadoop via SAP Data Services 4.1
- Ability to load and read large volumes of data into and out of SAP HANA from/to Hadoop via SAP Data Services 4.1
- Reading from and loading into Hadoop
  - Hive Support (table-like access to Hadoop data)
  - Direct HSFS file support through PIG scripts
SAP Predictive Analysis

- Intuitively design complex predictive models
- Visualize, discover, and share hidden insights
- Unleash Big Data with SAP HANA’s power
Smart Meter Analytics

- Identify energy consumption pattern that can be used for Customer Segmentation
- Smart meter data volume is huge

Using **SAP HANA + PAL K-Means** algorithm

- Clustering of smart meter data → applying the k-means clustering to >20 million x 96-dimensional vectors
- High performance clustering computation
Customer Analytics
Customer Revenue Performance Management

What:
- **Combination** of financial and sales data to get visibility of profitability and margins of a customer, a customer deal or group
- **Margin visibility** – decomposition of margin to highlight areas of improvements
- **Comprehensive** customer revenue analysis
- **Call for action** – take focus customers and define initiatives to improve sales success

Simulation and Business Advice
- Perform what-if scenarios to see the impact of pricing and discount changes
- Simulations in real time to provide advice on cross/up sell and product bundling
- Predictions on customer behaviour and profitability for more effective segmentation and targeting
Customer Analytics
Predictive Customer Segmentation & Targeting

What:
Segmentation on a wider and more granular set of data not requiring SAP CRM
Tracking of successful implementation of initiatives out of segmentation
All levels of sales can use segmentation features to define own target customers to improve sales
Predictive segmentation by identifying common characteristics within a given group

Real-time simulation and segment building
- More granular targeting leveraging predictive analytics
- Improved alignment amongst Sales, Marketing and Product Development to focus on target Customers
Performance and Insight Optimization Services
Enabling complex use cases with SAP HANA

Examples of innovative business solutions advanced by SAP HANA:

- **Tax Compliance and discovery engine for public sector**

- **Affinity insight and forecasting for retail**

- **High performance compliance engine for chemical companies**

**Industry-specific optimization**
Algorithms and processes

**Industry-specific predictive modeling**
Proprietary models for retail, banking, utilities, manufacturing, and beyond

**Back end**
- Sophisticated aggregation and cleansing algorithms
- High volume industry specific data relationship algorithms

SAP HANA
## Startups Powered by SAP HANA

<table>
<thead>
<tr>
<th>Predictive and Analytics / Security</th>
<th>AlertEnterprise Security Convergence solutions combine Situational Intelligence with Command and Control capabilities for Critical Infrastructure Protection.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geospatial &amp; Visual Analytics</td>
<td>Next generation geospatial and visual analytics solutions transform massive volumes of real-time disparate data into intuitive visual displays. Space-Time Insight’s situational intelligence software helps visualize operations with absolute clarity, analyze the cause of a problem and determine how to prevent another occurrence, and act instantly to address any situation.</td>
</tr>
<tr>
<td>Predictive and Analytics / Financial</td>
<td>The Taulia Invoicement® Suite includes a Dynamic Discounting module and self-service vendor portal. Together, these tools reduce Accounts Payable support overhead, improve supplier relations and generate significant cash savings for corporations.</td>
</tr>
<tr>
<td>Social and Predictive Analytics</td>
<td>From campaigns to communities, #NextPrinciples empowers anyone to monitor, manage and act on social media interactions across channels.</td>
</tr>
</tbody>
</table>
Agenda

- The Predictive Analytics Landscape
- Predictive Analytics with SAP HANA
- Customers Benefit from Predictive Analytics with SAP HANA
- Predictive Applications for SAP HANA
- Demo
Companies can no longer focus solely on delivering the best product or service. To succeed, they must:

- Uncover hidden customer / employee / vendor / partner trends and insights
- Anticipate behavior and take proactive action
- Empower your team with intelligent next steps to exceed customer expectations
- Create new offers to increase market share and profitability
- Develop and execute a customer-centric strategy
- Target the right offers to the right customers through the best channels at the most opportune time
Key Learnings

- New market forces are changing the landscape and offering new opportunities

- Predictive analysis can be for everyone in the business

- SAP HANA is the cornerstone of SAP’s vision for predictive analytics
Thank you for participating.

Please provide feedback on this session by completing a short survey via the event mobile application.

SESSION CODE: 0212

Learn more year-round at www.asug.com