Accelerating SPSS Modeling and Data Mining using IBM DB2 Analytics Accelerator on z

Stefan Wirag
stefan.wirag@de.ibm.com

03/20/2013
Scope of this Session

- We’re talking about a feature proposal where multiple different products (DB2 for z/OS, DB2 Analytics Accelerator, SPSS Modeler) are working together to deliver new analytical capability to the System z platform.

- Not all of the here shown is available as product or is currently in product development.

- With the help of prototypes, we want to share current thoughts on possible integration scenarios without implying any specific product enhancement or solution.

- The goal of this session is to have an interactive discussion about customer requirements and feasible solutions.
Agenda

- Overview Modeling & Scoring

- Existing SPSS Modeling & Scoring on zEnterprise

- Prototype: SPSS Modeling accelerated by IBM DB2 Analytics Accelerator
Why Modeling & Scoring?

How to:

- Attract more and better customers
- Grow customer value
- Retain profitable customers
- Mitigate risk
- Detect and prevent fraud
Driving Smarter Business Outcome

Capture

Data Collection

Predict

Text Mining

Data Mining

Statistics

Platform

Act

Deployment Technologies

Pre-built Content

Attract

Up-sell

Retain
What is involved in modeling and scoring?

**Align**

**Analyses**
- Segments
- Profiles
- Scoring models
...

**Data:**
- Demographics
- Account activity
- Transactions
- Channel usage
- Service queries
- Renewals
...

**Scoring**

**Use those predicted variables to score data & identify the best possible future outcomes**

**Act**

**Identify predictive models/patterns** found in historical and transactional data
SPSS Stream Example: Fraud Prediction for Banking Blueprint

Model Creation:

Batch Scoring:
SPSS Analytics on System z: High-level Concept

All data under z/OS control
IBM DB2 Analytics Accelerator (IDAA) – Cost-effective Analytics

Building on the Core Values

**Fast**
- Complex queries run up to 2000x faster while retaining single record lookup speed

**Cost Saving**
- Eliminate costly query tuning while offloading complex query processing

**Appliance**
- No applications to change, just plug it in, load the data, and gain the value

Reducing the Cost of High Speed Analytics

**Choice of historical data location**
- Reduces host data warehouse storage usage by over 95%
- Significantly reduces cost of host storage resources and administration

**Real time analytics**
- Data changes are propagated for high-speed use as they occur
- Current information is available for right-time decisions
- Extends the accelerator use to reporting on operational data

**Faster data refresh**
- Data available for analytics faster
- Refresh of tables or partitions is much faster and more efficient
- More optimized version of unloading data from DB2

**New queries**
- More queries eligible for acceleration

**New capacity**
- Now expandable to 960 cores and 1.28 petabytes
Agenda

- Overview Modeling & Scoring

- Existing SPSS Modeling & Scoring on zEnterprise

- Prototype: SPSS Modeling accelerated by IBM DB2 Analytics Accelerator
IBM SPSS Modeler Logical Infrastructure with zEnterprise

- Improved availability and security
- Faster communication thru Hipersockets between DB and Modeler Server

IBM DB2 Analytics Accelerator (IDAA)

System EC12

- z/OS DB2
  - Scoring Adapter (UDF)
- SPSS Modeler Server
  - Data Access Package (ODBC)
- Linux on z
- z/VM
- Hipersocket communication

zBX

- SPSS Modeler Client
- Sentinel RMS License Manager
- Windows
- AIX | Windows
- IEDN communication

Model Specialists

SPSS Modeler Client
Stream/Model Creation with SPSS Modeler and DB2 for z/OS

- Data preparation for modeling accelerated by IDAA (Table acceleration needs to be specified in DB2)
- z/OS DB2 Data Sharing allows to isolate analytic workload from business-critical transactional workload
Deployment of built models to DB2 is triggered from SPSS Modeler Client.
Real-time In-Transactional Scoring with DB2 v10 for z/OS

- Delivers better, more profitable decisions, using the latest data, at the point of customer impact
  - Enables more informed customer interaction
  - Improves fraud identification and prevention

- With improved accuracy, speed and performance while reducing cost and complexity
  - Improves accuracy by scoring new and relevant data directly within the OLTP application
  - Scales to large data volumes to improve accuracy of data models
  - Delivers the performance needed to meet and exceed SLAs of OLTP applications
  - No need to move data real-time from z/OS to distributed
  - Minimizes demand on network, HW, SW, and resources
Agenda

- Overview Modeling & Scoring
- Existing SPSS Modeling & Scoring on zEnterprise
- Prototype: SPSS Modeling accelerated by IBM DB2 Analytics Accelerator
Prototype: SPSS Model Creation Accelerated by IDAA

- Data preparation & model build is processed and accelerated by IDAA
- Integration of model build and scoring
Prototype Results

- Prototype integrates into zEnterprise the whole life cycle for:
  - data analysis
  - modeling
  - model deployment

- Prototype shows:
  - IBM Netezza Analytics (INZA) covers most algorithms required to accelerate realistic fraud detection scenarios
  - IDAA is able to execute algorithms for predictive modeling highly integrated
  - Predictive models created by IDAA/Netezza can be exported back to z/OS DB2 and can be deployed in transactional contexts for scoring
  - Push down of analytics to IDAA/Netezza is significantly faster for some modeling algorithms
Prototype Results: Measured Performance Improvements using Netezza

Netezza/INZA n times faster than SPSS Modeler Server

![Chart showing performance improvements](chart.png)