Agile Deployment of Predictive Analytics Using Amazon EC2

Michael Zeller, Ph.D.
CEO Zementis, Inc.

EDM Summit
October 30, 2008
1. Predictive Analytics

2. Development, Integration, and Deployment of predictive models … from R to WS to PMML
   Focus on PMML, the Predictive Modeling Markup Language

3. Cloud Computing and Amazon EC2

4. Bringing it all together: from the desktop to the cloud
Predictive Analytics

- Science that makes decisions smarter by uncovering hidden data trends not obvious to the human eye.
- Predicts future customer behaviour with today's data.
Business Objective for Predictive Analytics

- Automate Decisions
- Make Smarter Decisions
- Ensure Consistent Decisions
- Improve Processes
- Agility to Quickly Change with Market Conditions

1. Leverage in Every Business Process
2. Shorten Time-to-Market
3. Reduce Complexity and Cost
Presentation Outline

1. Predictive Analytics

2. Development, Integration, and Deployment of predictive models
   ... from R to WS to PMML
   Focus on PMML, the Predictive Modeling Markup Language

3. Cloud Computing and Amazon EC2

4. Bringing it all together: from the desktop to the cloud
Development, Integration, and Deployment

**Development**
R allows for reliable data manipulation and model building

**Integration**
Web-service calls allow for fast integration

**Deployment**
PMML allows for easy expression and deployment of data transformations and data-mining models

Open Standards
Development, Integration, and Deployment

Development
R allows for reliable data manipulation and model building

Integration
Web-service calls allow for fast integration

Deployment
PMML allows for easy expression and deployment of data transformations and data-mining models
The R Project

- R is an integrated suite of software facilities for data manipulation, calculation and graphical display.
- R provides a wide variety of statistical techniques and is highly extensible.
- R is similar to the S language and environment developed at Bell Labs.
- It is Open Source and a GNU project.
- R is available for free at http://www.r-project.org/
## Integration

### Web-Services for Integration

- **Service Oriented Architecture (SOA)**
  - Defined as a group of services that communicate with each other
  - Implementation via Web Services
- Establishes a “loosely coupled” infrastructure
  - Allows the business to respond more quickly and cost-effectively to changes in market conditions
- Increases Interoperability
  - Integrate various systems
  - Agnostic to specific languages (Java, .Net, Cobol)
  - Utilize internal and external services
Deployment and Execution

- PMML is an **XML**-based language to
  - Define statistical and data mining models
  - Share models between compliant applications
- Standard for exchange of models to
  - Avoid proprietary issues and incompatibilities
  - Deploy models in operational infrastructure
- Clear separation of tasks
  - Model development vs. model deployment
  - Scientists focus on building the best model
  - Eliminates need for custom model deployment
- Ensures scalability and reliability
PMML Industry Support

Mature and Supported by Industry

- Data Mining Group [http://www.dmg.org](http://www.dmg.org)
- Mature standard
  - Current version 3.2
  - Active group and constant enhancements
- Vendor independent consortium
- Industry supporters
  - Major Players: IBM, Oracle, SAP, Microsoft
  - Analytics: SAS, SPSS, Fair Isaac, Zementis
  - Business Intelligence: MicroStrategy, Teradata
  - Open Source: R
**PMML**

Bringing data and Models Together

---

**Predictive Modeling Markup Language**

- A **Data Dictionary** defines all the raw data fields (including missing value strategy and outlier treatment).
- Several **Data Transformations** strategies allow for intelligent extraction of feature detectors from raw data (“data massaging”).
- A comprehensive list of **Data-Mining Models** offers power and flexibility.
- Post-processing of results allow for tailored decisions.

---

Data Transformations and Data-Mining Models come together in PMML.
Presentation Outline

1. Predictive Analytics

2. Development, Integration, and Deployment of predictive models … from R to WS to PMML
   Focus on PMML, the Predictive Modeling Markup Language

3. Cloud Computing and Amazon EC2

4. Bringing it all together: from the desktop to the cloud
Got Models…

What Now?

- Data Analysis
- Statistical Model
- PMML Export
Cloud Computing

Computing as a “Utility”
Evolve from “World Wide Web” to “World Wide Computer”

Applications/Services/Storage/CPUs on the Internet.
Incorporate via SOA Standards and Web Services.

Emerging Utility Providers: Amazon, Sun, Google...
Promise: Scalable, secure, reliable and more cost-effective than do-it-yourself.

You are already using it everyday:
Google search, Gmail, Salesforce.com, …
Cloud Computing Benefits

- Open Standards vs. Proprietary Code
- Select Best-of-Breed Services & Applications
- Avoid Vendor Lock-in

- Deployment in Minutes vs. Months
- No In-house Hardware/Software to Maintain
- Scales with Business Demand

- Operational Cost vs. Capital Expenditures
- No Long-term Commitment
- Only Pay for Actual Usage

SOA
Open Standards

Utility Computing Paradigm

SaaS
Pay-As-You-Go
Amazon Elastic Compute Cloud (EC2)

Cost-effective and Reliable
- Software as a Service (SaaS)
- Based on Amazon’s Infrastructure

Secure
- Dedicated, Controlled Instances
- HTTPS & WS-Security

Elastic
- Choice of Instance Type (S,L,XL)
- Launch Multiple Instances on Demand

Superior Time-to-Market
- Commission Your Instance in Minutes
- Minimal Learning Curve: Based on LINUX / Open Source and Commodity Hardware
Presentation Outline

1. Process overview and Predictive Analytics

2. Development, Integration, and Deployment of predictive models … from R to WS to PMML
   Focus on PMML, the Predictive Modeling Markup Language

3. Cloud Computing and Amazon EC2

4. Bringing it all together: from the desktop to the cloud
Bringing it all together

Recapping ...

Cloud Computing
- Internet-centric use of applications, services, or computers as a “utility”.
- Incorporates SaaS and SOA concepts.
- Scalable, Secure, Cost-effective.

Bridging the gap

Predictive Analytics
- Science that makes decisions smarter by uncovering hidden data trends not obvious to the human eye.
- Predicts future customer behaviour with today’s data.
From the Desktop to the Cloud: Bridging the Gap with ADAPA

Model Building
Extensive data analysis and manipulation as well as selection of the most appropriate statistical modeling technique. These steps can then be represented as a PMML file.

Model Execution
Amazon EC2 offers utility computing with virtually unlimited scalability as well as flexible choice of server size based on memory and processing needs.
SaaS Predictive Analytics on Amazon EC2
The ADAPA Example

**Scalable Execution Platform**
Data transformation and model execution in real-time or batch mode.

**Environment to Manage Models and Rules**
Deploys one or many models or rules sets. Manages and maintains these through a web console.

**Framework for SOA-based IT Integration**
Completely standards-based and easily integrated into any existing infrastructure.

**ADAPA is not ...**
Not a model development environment.
Bridging the Gap: Highlights

- Ability to Quickly Deploy Predictive Models
- Open Standards for Integration and Models
- Leverages a Scalable & Secure Infrastructure
- Ability to Execute Predictive Models in Real-time
- Web 2.0 Support
- Low TCO and Fast Time-to-Market
1 through 6 – From Raw Data to Smart Decisions

1. Data Extraction and Analysis
2. Model Building
3. PMML Export
4. PMML Import
5. Web-Service Calls
6. Model Execution
Fast and effective creation of your presentation

Appealing visualization of your contents

Your Enterprise Decision Management Strategy
Thank You!

E-mail: info@zementis.com

U.S.A

6125 Cornerstone Court East
Suite 250
San Diego, CA, 92121

Tel: +1 619 330-0780
Fax: +1 858 535-0227

Asia

19/F., Unit A
Ho Lee Commercial Building
38-44 D'Aguilar Street
Central, Hong Kong (S.A.R.)

Tel: +852 2868-0878
Fax: +852 2845-6027