The Oracle Data Mining Machine Bundle: Zero to Predictive Analytics in Two Weeks
Collaborate 15 IOUG

Presentation #730
Tim Vlamis and Dan Vlamis
Vlamis Software Solutions
816-781-2880
www.vlamis.com
Your comments, questions, opinions, and ideas are more important to this session than our prepared PowerPoint slides!
Vlamis Software Solutions

• Vlamis Software founded in 1992 in Kansas City, Missouri
• Developed more than 200 Oracle BI systems
• Specializes in ORACLE-based:
  • Data Warehousing
  • Business Intelligence
  • Data Mining and Predictive Analytics
  • Data Visualization
• Expert presenter at major Oracle conferences
• www.vlamis.com (blog, papers, newsletters, services)
• Co-authors of book “Data Visualization for OBI 11g”
• Co-author of book “Oracle Essbase & Oracle OLAP”
• Oracle University Partner
• Oracle Gold Partner
Tim Vlamis and Dan Vlamis

• Tim Vlamis
  • 25+ years experience in business modeling and valuation, forecasting, and scenario analyses
  • Instructor for Oracle University’s Data Mining Techniques and Oracle R Enterprise Essentials Courses
  • Professional Certified Marketer (PCM) from AMA
  • Adjunct Professor of Business Benedictine College
  • MBA Kellogg School of Management (Northwestern University)
  • BA Economics Yale University

• Dan Vlamis
  • Founded Vlamis Software Solutions in 1993
  • 25+ years in business intelligence, dimensional modeling
  • Oracle Ace Director
  • Developer for IRI (expert in Oracle OLAP and related)
  • BA Computer Science Brown University

Copyright © 2015, Vlamis Software Solutions, Inc.
Top Mistakes

• Bringing in experts who are incented to prove that predictive analytics is hard.
• Choosing a starting project involving a complex and poorly understood business process.
• Choosing a starting project based solely on “interesting data”.
• Setting expectations that the first project will reveal previously unknown insights worth millions.
• Designing an enormous project.
• Feeding the “Golden Fleece Award” mentality.
Think about results

• Results drive analytic projects.
• Present results quickly.
• Get executives used to iterative, developmental analytics projects.
• Build staff capabilities and set them up for success.
• Keep results flowing and avoid major transitional hurdles.
• Plan for success, plan for production.
Oracle Data Mining

- Oracle Data Mining is an option for the Enterprise Edition of the Oracle Database.
- A collection of APIs and specialized SQL functions.
- Includes a large number of specialized algorithms and built-in procedures.
- Makes use of many built-in capabilities of the Oracle Database
- ODM typically refers to “Oracle Data Mining”
• Ranking functions
  • rank, dense_rank, cume_dist, percent_rank, ntile

• Window Aggregate functions (moving and cumulative)
  • Avg, sum, min, max, count, variance, stddev, first_value, last_value

• LAG/LEAD functions
  • Direct inter-row reference using offsets

• Reporting Aggregate functions
  • Sum, avg, min, max, variance, stddev, count, ratio_to_report

• Statistical Aggregates
  • Correlation, linear regression family, covariance

• Linear regression
  • Fitting of an ordinary-least-squares regression line to a set of number pairs.
  • Frequently combined with the COVAR_POP, COVAR_SAMP, and CORR functions

Descriptive Statistics
• DBMS_STAT_FUNCS: summarizes numerical columns of a table and returns count, min, max, range, mean, median, stats_mode, variance, standard deviation, quantile values, +/- n sigma values, top/bottom 5 values

• Correlations
  • Pearson’s correlation coefficients, Spearman’s and Kendall's (both nonparametric).

• Cross Tabs
  • Enhanced with % statistics: chi squared, phi coefficient, Cramer's V, contingency coefficient, Cohen's kappa

• Hypothesis Testing
  • Student t-test, F-test, Binomial test, Wilcoxon Signed Ranks test, Chi-square, Mann Whitney test, Kolmogorov-Smirnov test, One-way ANOVA

• Distribution Fitting
  • Kolmogorov-Smirnov Test, Anderson-Darling Test, Chi-Squared Test, Normal, Uniform, Weibull, Exponential
Oracle Data Miner

- Oracle Data Miner is a front end GUI for Oracle Data Mining.
- Extension for Oracle SQL Developer 3.x, 4.x a free utility program from Oracle that facilitates interaction with databases.
- Functions as an object oriented programming interface for designing data mining processes and procedures.
### Oracle Data Mining Algorithms

<table>
<thead>
<tr>
<th>Problem</th>
<th>Algorithm</th>
<th>Applicability</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Classification</strong></td>
<td>Logistic Regression (GLM)</td>
<td>Classical statistical technique</td>
</tr>
<tr>
<td></td>
<td>Decision Trees</td>
<td>Popular / Rules / transparency</td>
</tr>
<tr>
<td></td>
<td>Naïve Bayes</td>
<td>Embedded app</td>
</tr>
<tr>
<td></td>
<td>Support Vector Machine</td>
<td>Wide / narrow data / text</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Regression</strong></td>
<td>Linear Regression (GLM)</td>
<td>Classical statistical technique</td>
</tr>
<tr>
<td></td>
<td>Support Vector Machine</td>
<td>Wide / narrow data / text</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Anomaly Detection</strong></td>
<td>One Class SVM</td>
<td>Unknown fraud cases or anomalies</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Attribute Importance</strong></td>
<td>Minimum Description Length</td>
<td>Attribute reduction</td>
</tr>
<tr>
<td></td>
<td>Principal Component Analysis</td>
<td>Identify useful data</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Reduce data noise</td>
</tr>
<tr>
<td><strong>Association Rules</strong></td>
<td>Apriori</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Market basket analysis</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Next Best Offer</td>
</tr>
<tr>
<td><strong>Clustering</strong></td>
<td>Hierarchical K-Means</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Hierarchical O-Cluster</td>
<td>Product grouping</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Text mining</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Gene and protein analysis</td>
</tr>
<tr>
<td><strong>Feature Extraction</strong></td>
<td>Nonnegative Matrix Factorization (NMF)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Singular Value Decomposition (SVD)</td>
<td>Text analysis</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Feature reduction</td>
</tr>
</tbody>
</table>

---

Copyright © 2015, Vlamis Software Solutions, Inc.
ODM Machine Bundle Overview

- **Hardware**
  - Oracle Database Appliance

- **Software**
  - Oracle Database 12c or 11g (with options)
  - Oracle Advanced Analytics Option including Oracle Data Mining and Oracle R Enterprise
  - Oracle SQL Developer: Data Miner Add-in (free download)

- **Services**
  - Implementation and configuration from Vlamis Software Solutions (Oracle Gold Partner)
  - Oracle University Oracle Data Mining Techniques course (taught by Vlamis Software Solutions)
  - Market Basket Analysis Project performed on company data

- **Time frame:** 9 business days (less than 2 weeks)
Oracle Database Appliance

- A scalable engineered system
- Fast installation and configuration
- Enterprise level components
2 x 1RU x86 Servers. Each Server Contains:
- 2 x 12-core Intel Xeon Processors E5-2697 v2 processors
- 256 GB RAM (16 x 16 GB)
- 2 x 600 GB Boot Disks (mirrored)
- 4 x 10GbE Ports
- Redundant 10GbE Interconnect

1 x 2RU Storage Shelf – Direct-Attached
- 4 x 2.5" 200 GB SSDs for Redo
- 20 x 2.5” 900 GB HDDs for Data
Oracle Database Appliance X4-2
Storage Expansion Shelf Specifications

Double Available Storage Capacity

- Support for Additional 1 x 2RU Storage Expansion Shelf
  - Direct-Attached
  - 4 x 2.5” 200 GB SSDs for Redo
  - 20 x 2.5” 900 GB HDDs for Data
Oracle Database Appliance X5-2

• 2 Servers, each with
  – 36 Intel Xeon CPU cores
  – 256 GB memory, expandable to 768 GB
  – Redundant InfiniBand Interconnect
  – 10 GbE Public Network

• Storage Shelf
  – 800 GB flash log storage
  – 1.6 TB flash cache storage
  – 64 TB HDD
ODA X5-2 Storage Expansion Shelf

Zero-Admin/Online Storage Expansion

• Double available storage capacity
  – Additional 64 TB HDD, 128 TB total for DATA
  – Additional 800 GB SSD, 1.6 TB total for REDO
  – Additional 1.6 TB SSD, 3.2 TB total for FLASH

• Zero administration
  – Automatically integrates when plugged in
  – Data automatically distributes to new shelf

• Online expand storage
  – Hot-plug storage expansion shelf
  – No database downtime
Oracle Data Mining Training (2 days)

- Introduction
- Data Mining Concepts and Terminology
- The Data Mining Process
- Introducing Oracle Data Miner 11g Release 2
- Using Classification Models
- Using Regression Models
- Using Clustering Models
- Performing Market Basket Analysis
- Performing Anomaly Detection
- Deploying Data Mining Results
Schedule

- **Day 1:**
  - Two consultants meet with client team to review project plan, review data sources, identification of best data to start with, set technical objectives for project

- **Day 2:**
  - Consultant One: Install ODA and configure to network (need support from client tech staff)
  - Consultant Two: Conduct first day of ODM class with client team

- **Day 3:**
  - Consultant One: Install new pluggable Database, SQL Developer
  - Consultant Two: Conduct second day of ODM class with client team

- **Day 4:**
  - Two consultants establish data plan for project with client and import data

- **Day 5:**
  - Consultant One: Prepare tables for mining (add keys, new tables, transforms, etc.)
  - Consultant Two: Document data plan

- **Day 6:**
  - Consultant Two: Build market basket workflow

- **Day 7:**
  - Consultant Two: Conduct market basket analyses

- **Day 8:**
  - Consultant Two: Prepare presentation of findings from market basket analyses

- **Day 9:**
  - Consultant Two: Deliver presentation with client
Methodology

• Implement a highly scalable infrastructure
• Establish a common foundational understanding of data mining
• Demonstrate the Value of Analytics by Completing a Market Basket Project Immediately
Scalable Infrastructure

• Design with growth in mind
• Maintain a common environment
• Avoid recoding for production
Establish a Shared Data Mining Foundation

• Create a common understanding of data mining techniques, language, processes.
• Everyone should know what others know of.
• Ask for and show patience with experts.
Deliver Fast Results

- Demonstrate value immediately
- Slow is smooth, smooth is fast
- Pick the right starting project
- Success builds support
Start with Oracle Cloud Database

• Bundle includes two months of Oracle Database EE High Performance Virtual Image
• Two week onsite activity time frame
• Vlamis configures cloud instance for OAA
• Same schedule as with ODA
Advantages of Oracle Cloud Database

- Highly scalable
- Secure
- Ideal for sandbox development projects
- Works well as a Proof of Concept project before hardware commitment
- Perfect for organizations who want to move to the cloud
Three Reasons

- Nothing is more scalable for analytics than Oracle Database 12c running on Oracle engineered system hardware.
- Nothing is more important for 21st Century organizations than to add deep analytic capability and to generate returns from their operational, proprietary data sets.
- Nothing is more important to a project’s success than the key people involved.
Oracle Test Drive

- Free to try out Oracle BI, Advanced Analytics and Big Data
- Go to www.vlamis.com/td
- Runs off of Amazon AWS
- Hands-on Labs based on Collaborate 2012 HOLs
- Test Drives for:
  - Oracle BI
  - Oracle Advanced Analytics
  - Big Data
- Once signed up, you have private instance for 5 hours
- Available now

Copyright © 2015, Vlamis Software Solutions, Inc.
BIWA Summit 2016, Jan 26-28
Oracle HQ Conference Center

Business Intelligence, Warehousing and Analytics
and Spatial
IOUG Special Interest Group

www.biwasummit.org
Add business card to basket or fill out card
<table>
<thead>
<tr>
<th>Presenter</th>
<th>Session</th>
<th>Time</th>
<th>Location</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dan and Tim Vlamis</td>
<td>OAUG</td>
<td>Mon 12:45 – 1:45 PM South Seas D</td>
<td>Data Visualization for Oracle Business Intelligence 11g</td>
<td></td>
</tr>
<tr>
<td>Dan and Tim Vlamis</td>
<td>OAUG</td>
<td>Mon 3:15 – 4:15 PM Coral B</td>
<td>Designing an Analytics Strategy for the 21st Century</td>
<td></td>
</tr>
<tr>
<td>Dan and Tim Vlamis</td>
<td>IOUG</td>
<td>Tues 2:00 – 3:00 PM Jasmine E</td>
<td>Forecasting, Prediction Models, and Time Series Analysis with Database Analytics and OBIEE</td>
<td></td>
</tr>
<tr>
<td>Dan and Tim Vlamis</td>
<td>IOUG</td>
<td>Wed 3:15 – 4:15 PM Banyan D</td>
<td>The Oracle Data Mining Machine Bundle: Zero to Predictive Analytics in Two Weeks</td>
<td></td>
</tr>
<tr>
<td>Jon Clark</td>
<td>IOUG</td>
<td>Thurs 12:15 – 1:15 PM Reef F</td>
<td>Using Cloud technology for Oracle Database and Oracle BI Sandboxes and Training Environments</td>
<td></td>
</tr>
</tbody>
</table>
Thank You for Attending Session #730

Presenter Information
Tim Vlamis, Consultant
Dan Vlamis, President
Vlamis Software Solutions, Inc.
816-781-2880
tvlamis@vlamis.com
dvlamis@vlamis.com

For more information go to www.vlamis.com