# Analyzing Big Data

# Heartland OUG Spring Conference 2014

**Dan Vlamis**

**Vlamis Software Solutions**

**816-781-2880**

**http://www.vlamis.com**

# 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
  - Design and integrated BI and DW solutions
  - Training and mentoring
- Expert presenter at major Oracle conferences
- www.vlamis.com (blog, papers, newsletters, services)
- Developer for IRI (former owners of Oracle OLAP)
- Co-author of book "Oracle Essbase & Oracle OLAP"
- Beta tester for OBIEE 11g
- Reseller for Simba and Nokia map data for OBIEE
- Conference chair for BIWA Summit 2013, 2014
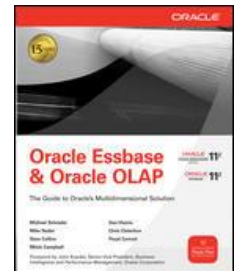
# Oracle Essbase & Oracle OLAP: The Guide to Oracle's Multidimensional Solution

- Published by Oracle Press
  - Dan Vlamis
  - Chris Claterbos
  - Michael Nader
  - David Collins
  - Floyd Conrad
  - Mitchell Campbell
  - Michael Schrader

- Covers both Oracle Essbase and Oracle OLAP
- 500 Pages

vlamis ™
SOFTWARE SOLUTIONS

# Analytical Options to Oracle Database

- Oracle OLAP
  - Defines a multi-dimensional data structure that allows information for highly complex calculations to done quickly.
  - Fast query performance and incremental update
  - Simplified access to analytic calculations
- Oracle Advanced Analytics (Data Mining & R)
  - Refers to the process of automatically sifting through data to find hidden patterns and make predictions.
  - Series of highly advanced algorithms and procedures.
  - Extends the "R" language to the Oracle Database
- Oracle Spatial
  - Provides the capability of relating data to geo positional coordinates, objects, and constructs.
  - Allows the construction and analysis of network topologies.

vlamis ™
SOFTWARE SOLUTIONS

# Spectrum of Oracle DB BI & Analytics

| OLAP | Data Mining & R | Spatial |
|------|-----------------|---------|
| Summaries, hierarchies and dimensional data | Knowledge discovery of hidden patterns | Spatial relationships between data |
| *"Analysis"* | *"Insight & Prediction"* | *"Location"* |
| **What is the** average **income** of mutual fund buyers, **by** region, **by** year? | **Who is likely to purchase a** mutual fund in the next 6 months and why? | **Where** were mutual funds purchased in the last 3 years? |

vlamis
SOFTWARE SOLUTIONS

# Competitive Advantage of BI & Analytics

**Optimization** — What's the best that can happen?

**Predictive Modeling** — What will happen next?

**Forecasting/Extrapolation** — What if these trends continue?

**Statistical Analysis** — Why is this happening?

**Alerts** — What actions are needed?

**Query/drill down** — Where exactly is the problem?

**Ad hoc reports** — How many, how often, where?

**Standard Reports** — What happened?

*Competitive Advantage*

*Degree of Intelligence*

ORACLE

**Analytic$**

**Access & Reporting**

**$$**

**Source:  Competing on Analytics, by T. Davenport & J. Harris**

vlamis
SOFTWARE SOLUTIONS

# Why OLAP for BI?

- BI often presents data dimensionally
- Dimensions are natural way to look at data
  - By, across, over, time, geography, product
  - Comparison of multiple dimension values
- Multi-dimensional storage of data speeds analysis
- Natural to express dimensional comparisons
  - Share of parent
  - Compared to last year
- Allows for hierarchical dimensions with multiple levels
  - E.g. by country, drill to state, drill to city

**vlamis** ™
**SOFTWARE SOLUTIONS**

# Oracle OLAP
## Leveraging Core Database Infrastructure

Oracle Database 11*g*

Data Warehousing

Warehouse Builder

OLAP

Data Mining
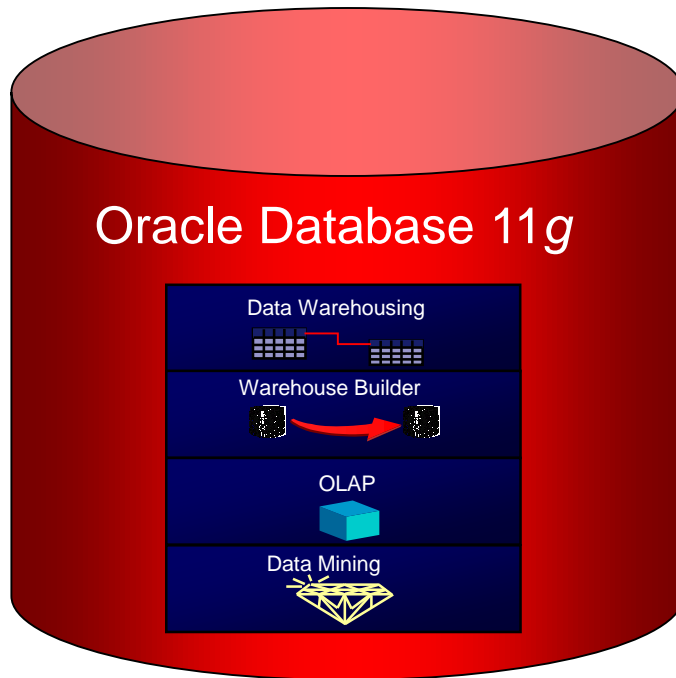
- Single RDBMS-MDBMS process
- Single data storage
- Single security model
- Single administration facility
- Grid-enabled
- Accessible by any SQL-based tool
- Embedded BI metadata
- Connects to all related Oracle data

vlamis
SOFTWARE SOLUTIONS

# Oracle OLAP

- A summary management solution for SQL based business intelligence applications
  - An alternative to table-based materialized views, offering improved query performance and fast, incremental update
- A full featured multidimensional OLAP server
  - Excellent query performance for ad-hoc / unpredictable query
  - Enhances the analytic content of Business intelligence application
  - Fast, incremental updates of data sets

vlamis
SOFTWARE SOLUTIONS

NY
25,000
customers

Los Angeles
35 customers

Precomputed

Computed when queried
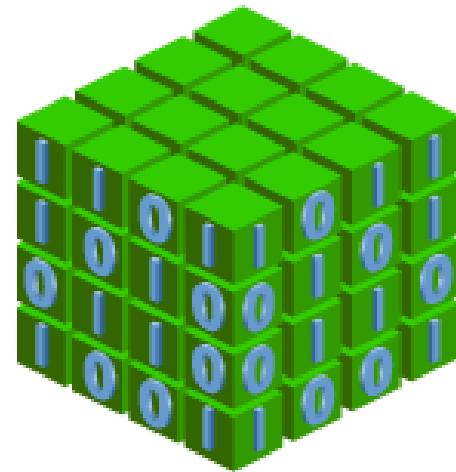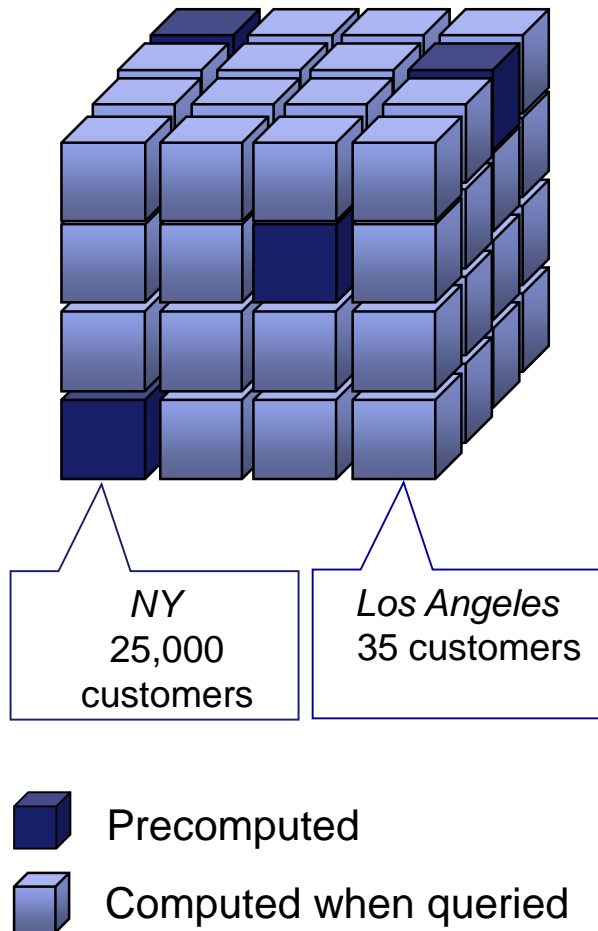
- Improves aggregation speed and storage consumption by pre-computing cells that are most expensive to calculate

- Easy to administer

- Simplifies SQL queries by presenting data as fully calculated

vlamis ™
SOFTWARE SOLUTIONS

# One Cube Accessed Many Ways…

- One cube can be used as
  - A summary management solution to SQL-based business intelligence applications as cube-organized materialized views
  - A analytically rich data source to SQL-based business intelligence applications as SQL cube-views
  - A full-featured multidimensional cube, servicing dimensionally oriented business intelligence applications

vlamis
SOFTWARE SOLUTIONS

# Cube Represented as Star Model
## Simplifies Access to Analytic Calculations

**PRODUCT_VIEW**
prod_id
subcategory
category
group

**CUSTOMER_VIEW**
cust_id
city
state
region

**SALES_CUBEVIEW**
day_id
prod_id
cust_id
chan_id
sales
profit
profit_yrago
profit_share_parent

**CHANNEL_VIEW**
chan_id
class
total

**TIME_VIEW**
day_id
quarter
month
year

SALES CUBE
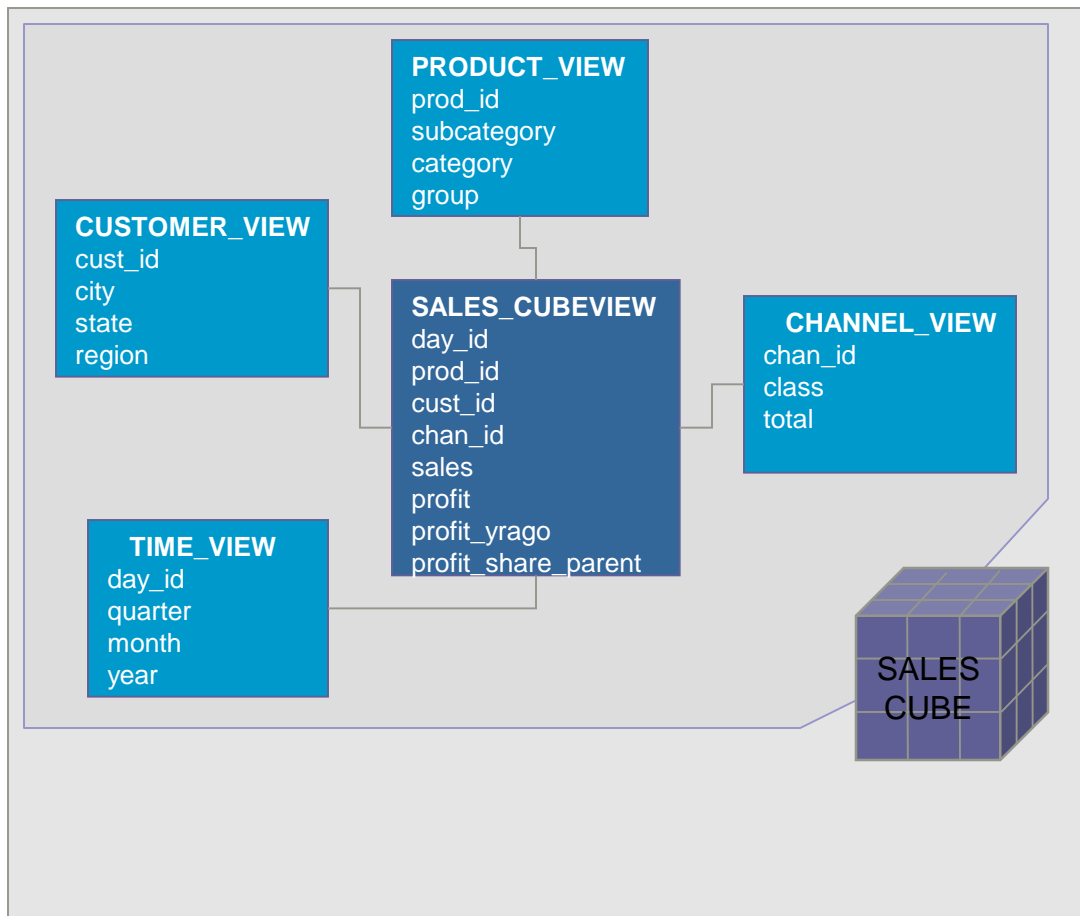
- Cube represented as a star schema
- Single cube view presents data as completely calculated
  - Analytic calculations presented as columns
  - Includes all summaries
- Automatically managed by OLAP

vlamis
**SOFTWARE SOLUTIONS**

# Essbase vs. Oracle OLAP

**Essbase**

- Separate server
- List price* $184K/CPU
- Separate admin
- Administer by LoB
- Must build cubes
- Part of middle tier
- Excellent writeback
- Query via MDX, XML/A

**Oracle OLAP**

- Built into Oracle DB
- List price* DB + $23K/CPU
- Admin same as Oracle DB
- Administer by IT
- Must build cubes
- Part of server tier
- Limited writeback
- Query via SQL (now MDX)

\* http://www.oracle.com/us/corporate/pricing/index.html

vlamis
SOFTWARE SOLUTIONS

# What is Data Mining?

- Automatically sifts through data to find hidden patterns, discover new insights, and make predictions

- Data Mining can provide valuable results:
  - Predict customer behavior *(Classification)*
  - Predict or estimate a value *(Regression)*
  - Segment a population *(Clustering)*
  - Identify factors more associated with a business problem *(Attribute Importance)*
  - Find profiles of targeted people or items *(Decision Trees)*
  - Determine important relationships and "market baskets" within the population *(Associations)*
  - Find fraudulent or "rare events" *(Anomaly Detection)*
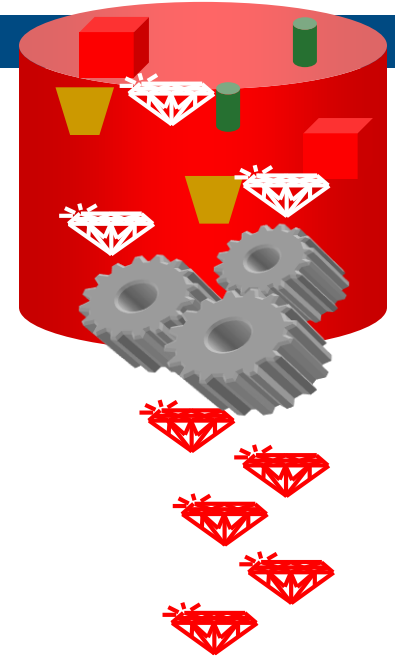
# In Database Data Mining
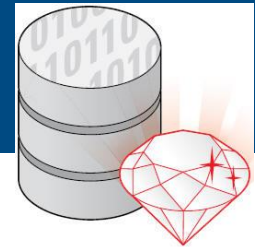
**Traditional Analytics**

**Oracle Data Mining**

| Traditional Analytics |
|---|
| **Data Import** |
| **Data Mining Model "Scoring"** |
| **Data Preparation and Transformation** |
| **Data Mining Model Building** |
| **Data Prep & Transformation** |
| **Data Extraction** |
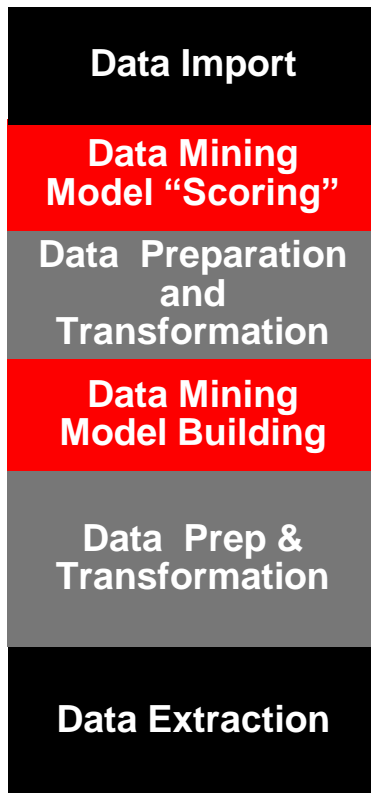
**$avings**

**Results**
- **Faster time for "Data" to "Insights"**
- **Lower TCO—Eliminates**
- **Data Movement**
- **Data Duplication**
- **Maintains Security**

Model "Scoring"
Data remains in the Database

Embedded data preparation

Cutting edge machine learning algorithms inside the SQL kernel of Database

SQL—Most powerful language for data preparation and transformation

Data remains in the Database

**Model "Scoring"**
**Embedded Data Prep**
**Model Building**
**Data Preparation**

**Hours, Days or Weeks**

| Source Data | Dataset s/ Work Area | Analytic al Process ing | Process Output | Target |
|---|---|---|---|---|
| | §sas | §sas | §sas | ORACLE |

**Secs. Mins or Hours**

ORACLE®
**Data Mining**

vlamis™
SOFTWARE SOLUTIONS

# 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"
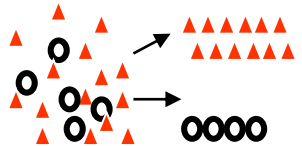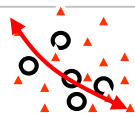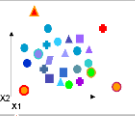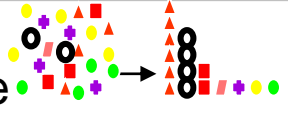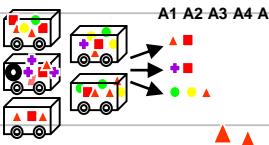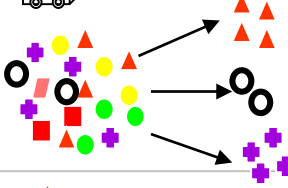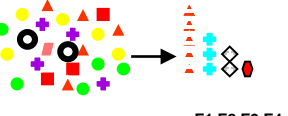
# Data Mining Provides
## *Better Information, Valuable Insights and Predictions*



**Cell Phone Churners** vs. *Loyal Customers*

**Income** (vertical axis)

**Customer Months** (horizontal axis)

Insight & Prediction

Segment #3:

IF CUST_MO > 7 AND INCOME < $175K, THEN Prediction = Cell Phone Churner, Confidence = 83%, Support = 6/39

Segment #1:

IF CUST_MO > 14 AND INCOME < $90K, THEN Prediction = Cell Phone Churner, Confidence = 100%, Support = 8/39

vlamis
SOFTWARE SOLUTIONS
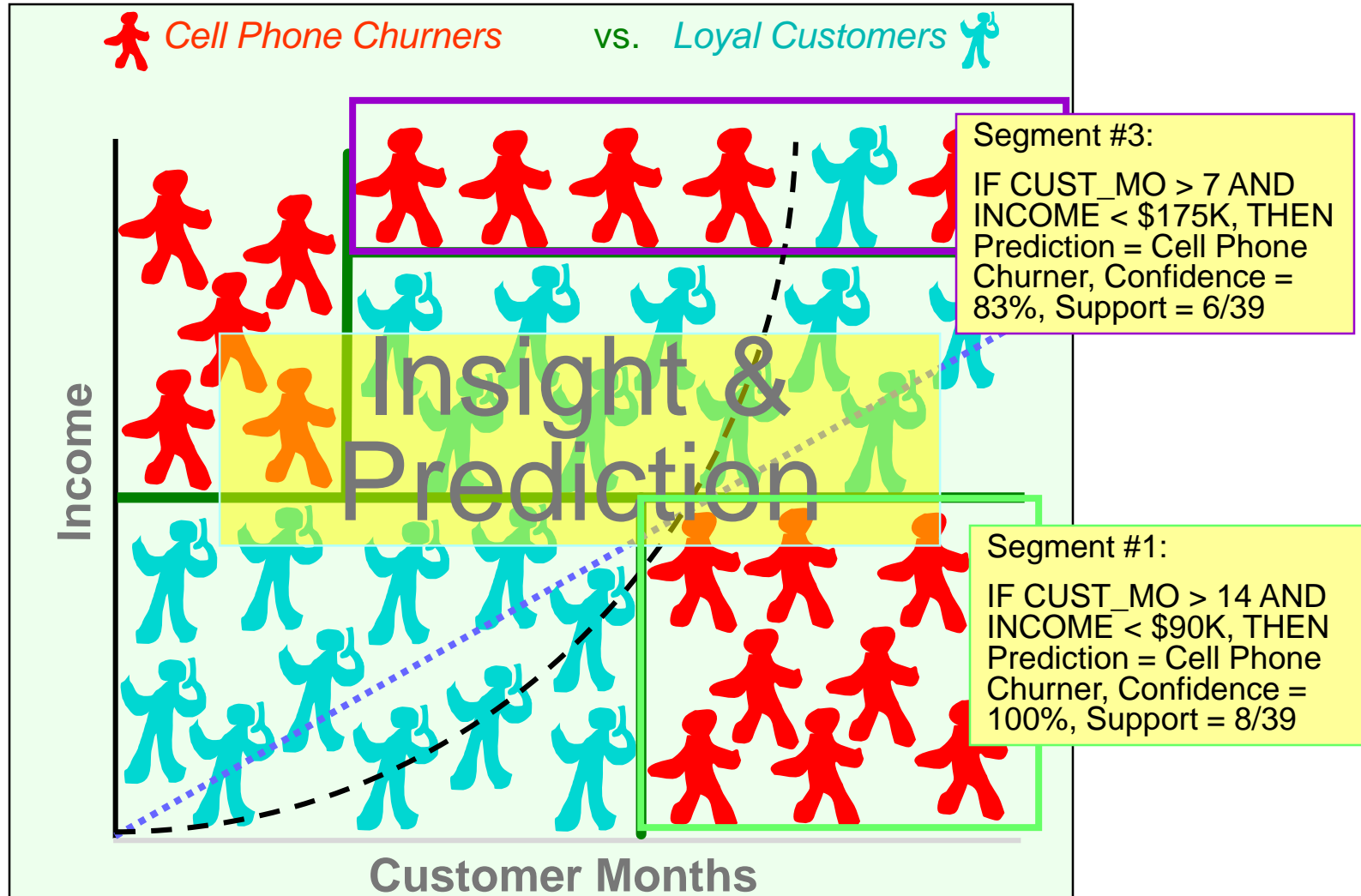
# Oracle Data Mining Algorithms

| Problem | Algorithm | Applicability |
|---|---|---|
| Classification | **Logistic Regression (GLM)** **Decision Trees** **Naïve Bayes** **Support Vector Machine** | **Classical statistical technique** **Popular / Rules / transparency** **Embedded app** **Wide / narrow data / text** |
| Regression | **Multiple Regression (GLM)** **Support Vector Machine** | **Classical statistical technique** **Wide / narrow data / text** |
| Anomaly Detection | **One Class SVM** | **Lack examples** |
| Attribute Importance | **Minimum Description Length (MDL)** | **Attribute reduction** **Identify useful data** **Reduce data noise** |
| Association Rules | **Apriori** | **Market basket analysis** **Link analysis** |
| Clustering | **Hierarchical K-Means** **Hierarchical O-Cluster** | **Product grouping** **Text mining** **Gene and protein analysis** |
| Feature Extraction | **NMF** | **Text analysis** **Feature reduction** |

vlamis ™
SOFTWARE SOLUTIONS

# 11g Statistics & SQL Analytics (Free)

- **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

vlamis ™
SOFTWARE SOLUTIONS

# Oracle Data Miner 11g
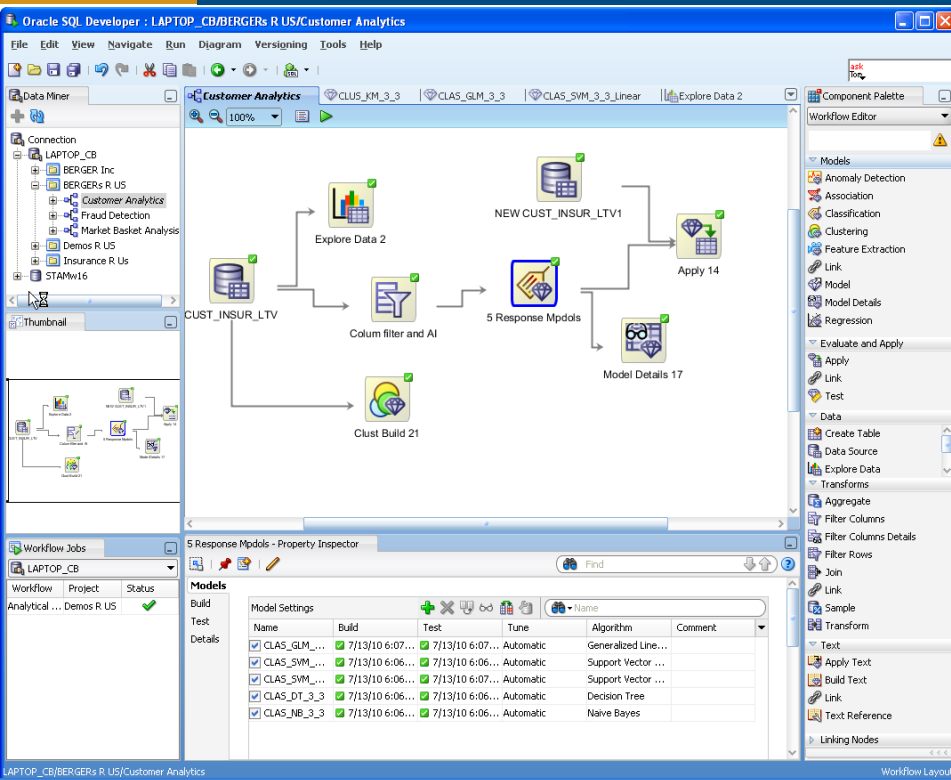


- Oracle Data Miner is a front end GUI for Oracle Data Mining.

- 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 programing interface for designing data mining processes and procedures.

vlamis
SOFTWARE SOLUTIONS

# Understand Model Details



- Interactive model viewers

# Oracle Data Mining & OBI 11g

ODM's predictions & probabilities are available in the Database for reporting using Oracle BI EE and other tools

**8.4 Oracle Datamining**

Home | Catalog | Dashboards ˅ | New ˅ | Open ˅ | Signed In As

| LTV Prediction | LTV Details | **Classification Tree** | LTV Probabilities | What If Scoring | Geo LTV P... | Map |

**Classification Tree**

⊞ Page Information (click to collapse or expand)

**Classification Tree**
Time run: 12/9/2011 1:03:03 PM

20 Actual Unit Price

| | | 2008 | 2009 | 2010 | Grand Total |
|---|---|---|---|---|---|
| ⊟ 0 - All Individuals | MEDIUM | 9.302 | 9.302 | 9.382 | **9.331** |
| ⊟ 1 - M_MARITAL_ST in 'DIVORCED', 'SINGLE' | MEDIUM | 9.207 | 9.329 | ✔ 9.421 | **9.322** |
| ⊟ 2 - M_CRDT_RATE <= 657.5 | LOW | 9.225 | 9.164 | 9.377 | **9.261** |
| 12 - M_INCOME_LVL in 'LEVEL 5', 'LEVEL 6', 'LEVEL 7', 'LEVEL 8', 'LEVEL 9' | MEDIUM | ✖ 8.904 | 9.131 | ✔ 9.670 | **9.261** |
| 13 - M_INCOME_LVL in 'LEVEL 1', 'LEVEL 2', 'LEVEL 3', 'LEVEL 4' | LOW | 9.345 | 9.176 | 9.259 | **9.261** |
| ⊟ 3 - M_CRDT_RATE > 657.5 | MEDIUM | | 9.193 | ✔ 9.462 | ✔ 9.454 | **9.370** |
| 14 - M_MONTHS_CONTACT <= 12.5 | VERY HIGH | ✖ 8.815 | ✔ 9.418 | ✖ 8.690 | **8.951** |
| ⊞ 4 - M_MONTHS_CONTACT > 12.5 | MEDIUM | 9.242 | ✔ 9.468 | ✔ 9.543 | **9.421** |
| ⊞ 7 - M_MARITAL_ST in 'MARRIED', 'WIDOW' | HIGH | 9.397 | 9.276 | 9.343 | **9.341** |

1- Revenue

| | | 2008 | 2009 | 2010 | Grand Total |
|---|---|---|---|---|---|
| ⊟ 0 - All Individuals | MEDIUM | 16,500,000 | 15,000,000 | 18,500,000 | **50,000,000** |
| ⊟ 1 - M_MARITAL_ST in 'DIVORCED', 'SINGLE' | MEDIUM | 8,155,247 | 7,589,505 | 9,289,014 | **25,033,766** |
| ⊟ 2 - M_CRDT_RATE <= 657.5 | LOW | 3,560,875 | 3,340,550 | 4,015,646 | **10,917,071** |
| 12 - M_INCOME_LVL in 'LEVEL 5', 'LEVEL 6', 'LEVEL 7', 'LEVEL 8', 'LEVEL 9' | MEDIUM | 938,983 | 889,059 | 1,189,016 | **3,017,058** |
| 13 - M_INCOME_LVL in 'LEVEL 1', 'LEVEL 2', 'LEVEL 3', 'LEVEL 4' | LOW | 2,621,892 | 2,451,491 | 2,826,630 | **7,900,013** |

...ails
...3:03 PM

| # | M23 Full Rule | Predicted LTV | # of Cust | 1- Revenue | Trend |
|---|---|---|---|---|---|
| 12 | M_MARITAL_ST in 'DIVORCED', 'SINGLE' ; AND M_CRDT_RATE <= 657.5; AND M_INCOME_LVL in 'LEVEL 5', 'LEVEL 6', | ◆ MEDIUM | 0 | | |
| 13 | M_MARITAL_ST in 'DIVORCED', 'SINGLE' ; AND M_CRDT_RATE <= 657.5; AND M_INCOME_LVL in 'LEVEL 1', 'LEVEL 2', | ▣ LOW | 0 | | |
| 14 | M_MARITAL_ST in 'DIVORCED', 'SINGLE' ; AND M_CRDT_RATE > 657.5; AND M_MONTHS_CONTACT <= 12.5 | △ VERY HIGH | 0 | | |
| 15 | M_MARITAL_ST in 'DIVORCED', 'SINGLE' ; AND M_CRDT_RATE > 657.5; AND M_MONTHS_CONTACT > 12.5; | ◆ MEDIUM | 0 | | |
| 16 | M_MARITAL_ST in 'DIVORCED', 'SINGLE' ; AND M_CRDT_RATE > 657.5; AND M_MONTHS_CONTACT > 12.5; | ▣ LOW | 0 | | |
| 17 | M_MARITAL_ST in 'DIVORCED', 'SINGLE' ; AND M_CRDT_RATE > 657.5; AND M_MONTHS_CONTACT > 12.5; | ◆ MEDIUM | 0 | | |
| 18 | M_MARITAL_ST in 'DIVORCED', 'SINGLE' ; AND M_MONTHS_CONTACT > 12.5; AND M_INCOME_LVL in 'LEVEL 1', 'LEVEL 2', | ◠ HIGH | 18 | 48,866 | |
| 19 | M_MARITAL_ST in 'MARRIED', 'WIDOW' ; AND M_INCOME_LVL in 'LEVEL 1', 'LEVEL 2', 'LEVEL 3', 'LEVEL 4' ; | ◆ MEDIUM | 0 | | |
| 20 | M_MARITAL_ST in 'MARRIED', 'WIDOW' ; AND M_INCOME_LVL in 'LEVEL 1', 'LEVEL 2', | ◠ HIGH | 0 | | |

vlamis ™
SOFTWARE SOLUTIONS

# Oracle Data Mining & Spatial



Customer "most likely" be
be `HIGH` and `VERY HIGH`
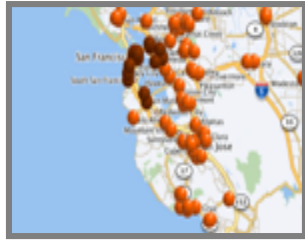value customer in the future

# What is Spatial Data?

- Business data that contains or describes location
  - Street and postal address (customers, stores, factory, etc.)
  - Sales data (sales territory, customer registration, etc.)
  - Assets (cell towers, pipe lines, electrical transformers, etc.)
  - Geographic features (roads, rivers, parks, etc.)
- Anything connected to a physical location
- Any data sets that contain "link and node" relationships between data objects. Can be directional or non-directional.
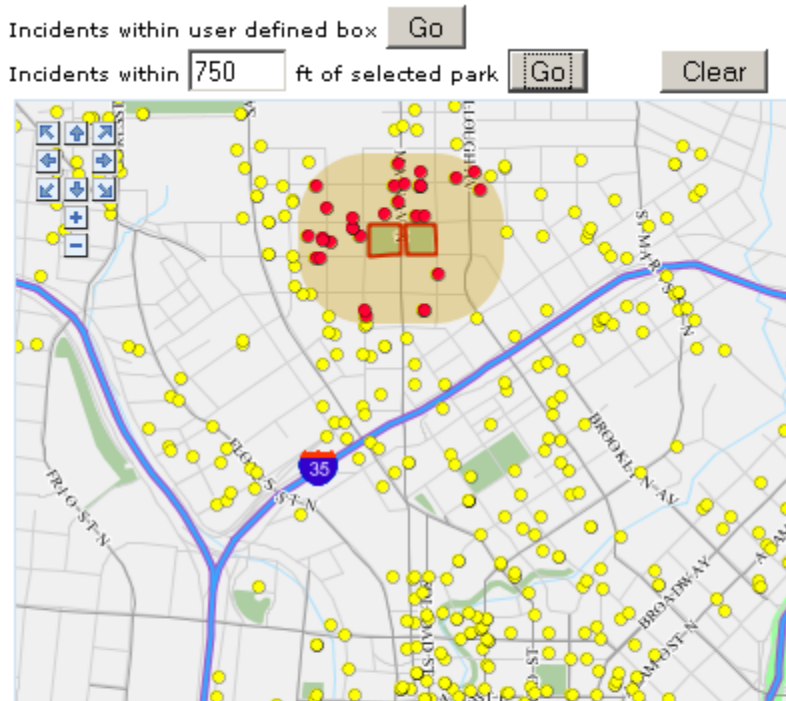
vlamis™

SOFTWARE SOLUTIONS

# Natively Manage All Geospatial Data



"Points"

"Lines"
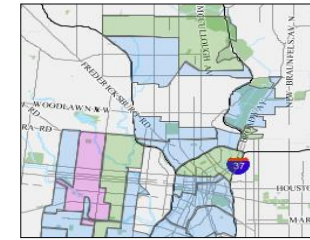
"Polygons"

Web Services (OGC)

Oracle database

**ORACLE** SPATIAL

Geocoding Routing

Rasters

Networks

Topologies

3D

vlamis TM

SOFTWARE SOLUTIONS
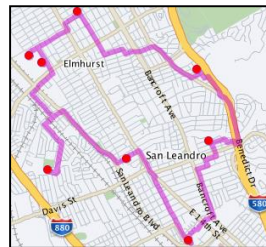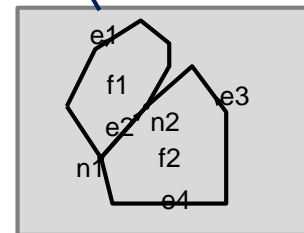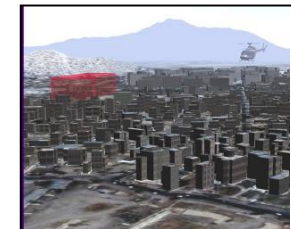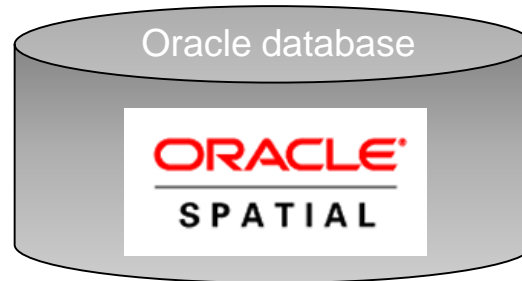
# Oracle Spatial Technologies

- **Oracle Locator**: Feature of Oracle Database XE, SE, EE

- **Oracle Spatial**: Priced option to Oracle Database EE

- **MapViewer**: Java application and map rendering feature of Oracle Fusion Middleware

- **Workspace Manager**: Long transactions feature of Oracle Database SE, EE

- **Bundled Map Content**: Major roads, administrative boundaries (city, county, state, country) - worldwide coverage from Navteq
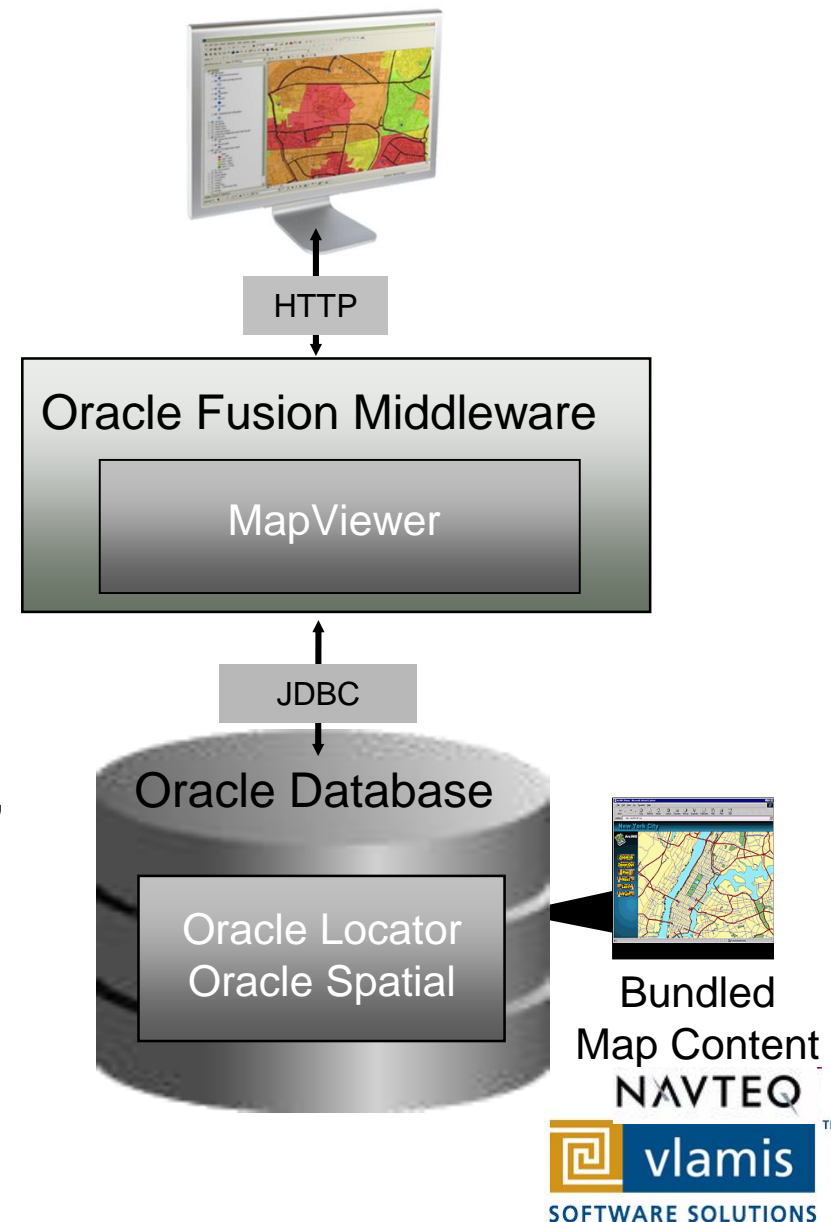
HTTP

**Oracle Fusion Middleware**

MapViewer

JDBC

**Oracle Database**

Oracle Locator
Oracle Spatial

Bundled Map Content

NAVTEQ

vlamis ™
SOFTWARE SOLUTIONS

# What is Spatial Data?

- Business data that contains or describes location
  - Street and postal address (customers, stores, factory, etc.)
  - Sales data (sales territory, customer registration, etc.)
  - Assets (cell towers, pipe lines, electrical transformers, etc.)
  - Geographic features (roads, rivers, parks, etc.)
- Anything connected to a physical location
- Any data sets that contain "link and node" relationships between data objects. Can be directional or non-directional.

# Depict and Detect Spatial Relationships
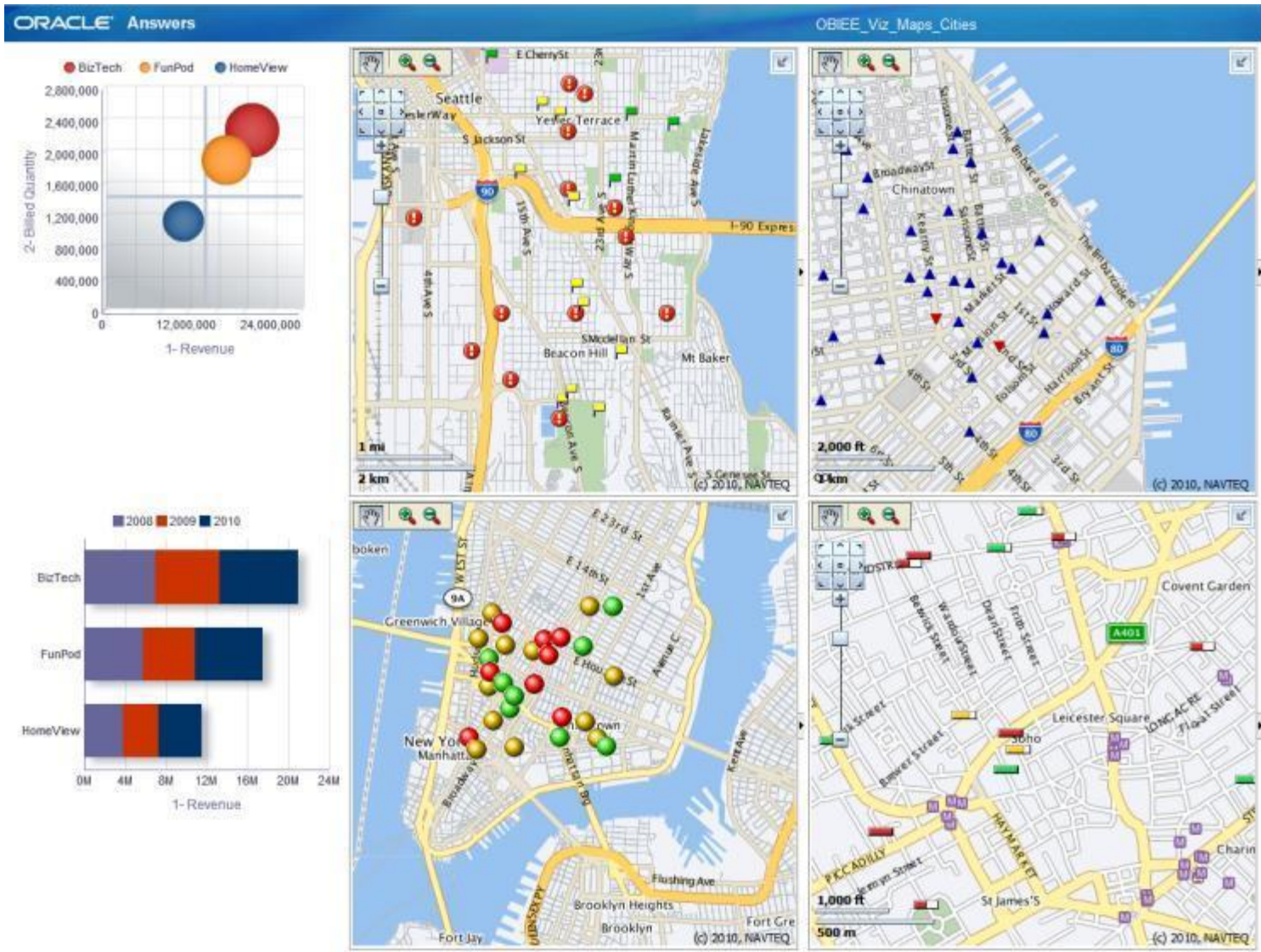
# Why Maps are Powerful

**Maps convey dense, multi-dimensional relationships in data faster and more intuitively than any other graphical display methodology.**

# Some Analysis Is Possible Only with Spatial Analytics

**Incidents within user defined box** [Go]

**Incidents within** [750] **ft of selected park** [Go] [Clear]



Show incidents within 750 ft
of selected park

## Complaint Detail

| Offense Desc | PD Desc | Date Key | Complaint Key | Service Area | Region |
|---|---|---|---|---|---|
| CRIMINAL MISCHIEF & RELATED OF | MISCHIEF, CRIMINAL 4, OF MOTOR | 18-Feb-03 | 1026 | 28 | Central |
| DANGEROUS DRUGS | CONTROLLED SUBSTANCE, POSSESSI | 10-Nov-02 | 30099 | 28 | Central |
| | | 10-Mar-03 | 40099 | 28 | Central |
| HARRASSMENT 2 | HARASSMENT,SUBD 1,CIVILIAN | 02-Aug-03 | 1064 | 32 | Central |
| | HARASSMENT,SUBD 3,4,5 | 04-Mar-03 | 1027 | 28 | Central |
| | | 04-May-03 | 31027 | 28 | Central |
| | | 04-Sep-03 | 41027 | 28 | Central |
| | | 19-Sep-03 | 41028 | 28 | Central |
| ROBBERY | ROBBERY,UNCLASSIFIED,OPEN AREA | 09-Jan-04 | 41032 | 28 | Central |

vlamis
SOFTWARE SOLUTIONS

# Oracle Locator and Oracle Spatial

- Oracle Locator is a **feature** of both Oracle Standard and Enterprise Database Editions.
- Oracle Locator provides basic location functionality.
  - Point, line, and polygon spatial locations (SDO_GEOMETRY)
  - Spatial indexing
  - Spatial operators that use the spatial index for performing spatial inquiries.

- Oracle Spatial is an **option** for Oracle Database Enterprise Edition
  - Provides extensive support for advanced spatial processing and analytics including routing, vector and raster data, topology and network models, and more.

# What is R?

- R is an Open Source scripting language and environment for statistical computing and graphics http://www.R-project.org/

- Popular alternative to SAS, SPSS & other proprietary statistical environments

- Around 2 million R users worldwide

- Thousands of R packages available

| | |
|---|---|
| Bayesian | Bayesian Inference |
| ChemPhys | Chemometrics and Computational Physics |
| ClinicalTrials | Clinical Trial Design, Monitoring, and Analysis |
| Cluster | Cluster Analysis & Finite Mixture Models |
| DifferentialEquations | Differential Equations |
| Distributions | Probability Distributions |
| Econometrics | Computational Econometrics |
| Environmetrics | Analysis of Ecological and Environmental Data |
| ExperimentalDesign | Design of Experiments (DoE) & Analysis of Experimental Data |
| Finance | Empirical Finance |
| Genetics | Statistical Genetics |
| Graphics | Graphic Displays & Dynamic Graphics & Graphic Devices & Visualization |
| HighPerformanceComputing | High-Performance and Parallel Computing with R |
| MachineLearning | Machine Learning & Statistical Learning |
| MedicalImaging | Medical Image Analysis |
| MetaAnalysis | Meta-Analysis |
| Multivariate | Multivariate Statistics |
| NaturalLanguageProcessing | Natural Language Processing |
| OfficialStatistics | Official Statistics & Survey Methodology |
| Optimization | Optimization and Mathematical Programming |
| Pharmacokinetics | Analysis of Pharmacokinetic Data |
| Phylogenetics | Phylogenetics, Especially Comparative Methods |
| Psychometrics | Psychometric Models and Methods |
| ReproducibleResearch | Reproducible Research |
| Robust | Robust Statistical Methods |
| SocialSciences | Statistics for the Social Sciences |
| Spatial | Analysis of Spatial Data |
| SpatioTemporal | Handling and Analyzing Spatio-Temporal Data |
| Survival | Survival Analysis |
| TimeSeries | Time Series Analysis |
| gR | gRaphical Models in R |

vlamis ™
SOFTWARE SOLUTIONS

# 2013 Rexer Analytics Data Miner Survey:

- 6th survey since 2007

- 68 questions

- 10,000+ invitations emailed, plus promoted by newsgroups, vendors, and bloggers

- Respondents: 1,259 data miners from 75 countries

- Data collected in first half of 2013
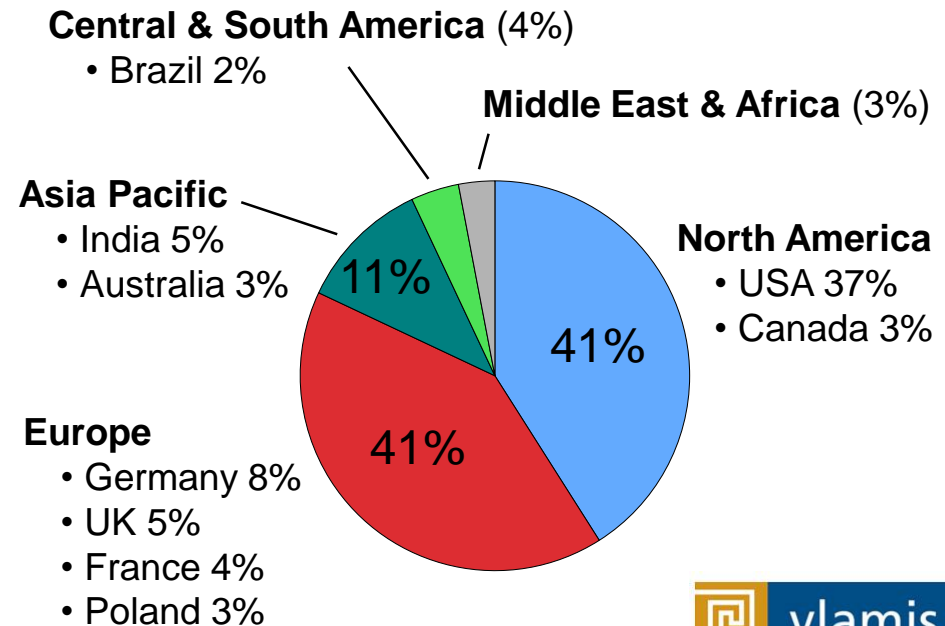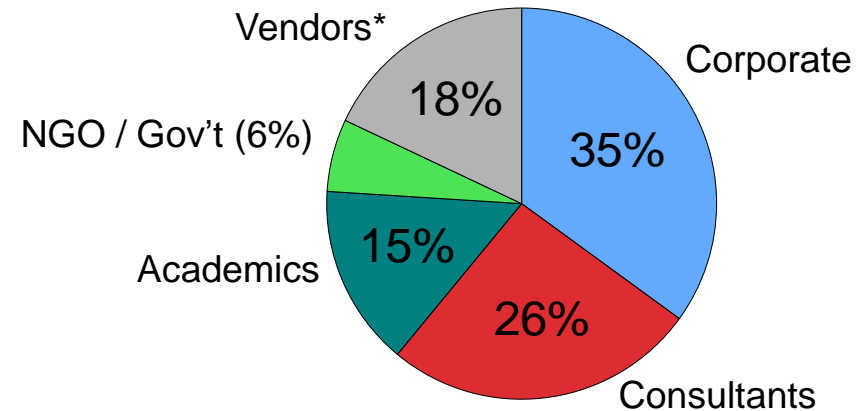
*Data from software vendors is excluded from analyses in this presentation unless otherwise noted.

**Pie chart 1:**
- Corporate 35%
- Consultants 26%
- Academics 15%
- NGO / Gov't (6%)
- Vendors* 18%

**Pie chart 2:**

Central & South America (4%)
- Brazil 2%

Asia Pacific
- India 5%
- Australia 3%

Middle East & Africa (3%)

North America 41%
- USA 37%
- Canada 3%

Europe 41%
- Germany 8%
- UK 5%
- France 4%
- Poland 3%

11%

vlamis
SOFTWARE SOLUTIONS

The proportion of data miners using R is rapidly growing, and since 2010, R has been the most-used data mining tool.  While R is frequently used along with other tools, an increasing number of data miners also select R as their primary tool.  Among data miners who say they are likely to switch their primary package in the coming year, R is frequently identified as the tool they are plan to switch to – more than 2.5 times more often that any other tool.

## R Usage



70% of data miners report using R

24% of data miners select R as their primary tool
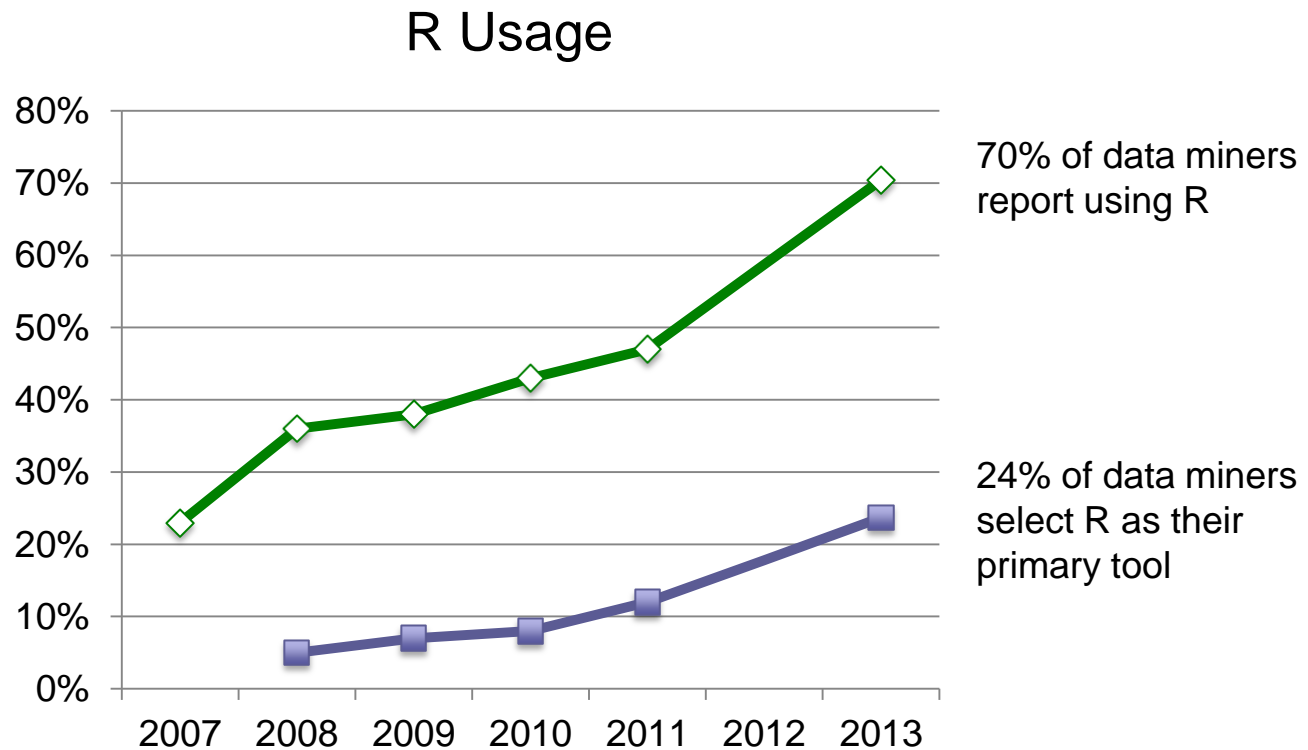
# Oracle R Enterprise

- Part of the Advanced Analytics Option to the Oracle Database Enterprise Edition
- Provides transparent access to database-resident data from R
- Embedded R script execution through database managed R engines with SQL language integration
- Provides data and task parallelism and full power of Oracle database for R
- Enables advanced statistics for in-database execution
- Integrates R into the IT software stack
- Extends and enhances open source R

vlamis
SOFTWARE SOLUTIONS

# Oracle R Distribution

*ability to dynamically load:*
Intel Math Kernel Library (MKL)
AMD Core Math Library
Solaris Sun Performance Library

**+**

Oracle Support

- Improved scalability at client and database for embedded R execution
- Enhanced linear algebra performance using Intel's MKL, AMD's ACML, and Solaris Sun Performance Library
- Enterprise support for customers of Oracle Advanced Analytics option, Big Data Appliance, and Oracle Linux
- Available as a free download from Oracle
- Oracle to contribute bug fixes and enhancements to open source R

vlamis
SOFTWARE SOLUTIONS

# Other R Offerings

- ROracle
  - Open source Oracle database interface driver for R based on OCI
  - Maintained by Oracle, optimizations and bug fixes released to open source community
- Oracle R Connector for Hadoop
  - R interface to Oracle Hadoop Cluster on Big Data Appliance
  - Access and manipulate data in HDFS, database, and file system
  - Write MapReduce functions using R and execute through R
- Rstudio
  - Popular open source user interface for R
  - Integrated Development Environment
- Rcommander
  - Extended GUI for R
  - Used for plugins (e.g. text mining)
  - Often used with Rstudio

vlamis
SOFTWARE SOLUTIONS

# R now integrated into OBIEE 11.1.1.7

# R now integrated into OBIEE 11.1.1.7

# 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 sign up, you have private instance for 5 hours
- Available now

vlamis
SOFTWARE SOLUTIONS

# Big Data Test Drive at Vlamis.com/td

# Oracle Big Data Lite VM

# Oracle MoviePlex



## • Introduction

- Oracle MoviePlex is an on-line movie streaming company
- Like many other on-line stores, they needed a cost effective approach to tackle their "big data" challenges
- They recently implemented Oracle's Big Data Platform to better manage their business, identify key opportunities and enhance customer satisfaction

# Thank You!

Thank You for Attending **Analyzing Big Data**

Presenter Information

Dan Vlamis, President

Vlamis Software Solutions, Inc.

816-781-2880

dvlamis@vlamis.com

For more information go to www.vlamis.com

vlamis
SOFTWARE SOLUTIONS