Blazing BI with Oracle DB Analytical Options:
Oracle OLAP, Oracle Data Mining, and Oracle Spatial

Heartland OUG October 20, 2011

Dan Vlamis and Tim Vlamis
Vlamis Software Solutions
816-781-2880
http://www.vlamis.com

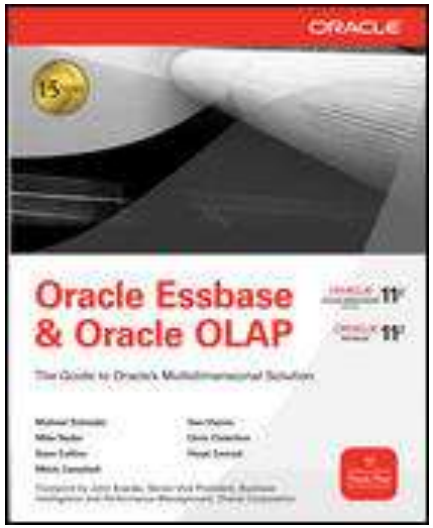
Vlamis Software Solutions, Inc.

- Founded in 1992 in Kansas City, Missouri
- Oracle Partner and reseller since 1995
- Developed more than 200 Oracle BI systems
- Specializes in ORACLE-based:
  - Data Warehousing
  - Business Intelligence
  - Data Transformation (ETL)
- Delivers
  - Design and integrated BI and DW solutions
  - Training and mentoring
- OBIEE 11g beta program participant
- Expert presenter at major Oracle conferences
- www.vlamis.com (blog, papers, newsletters, services)

vlamis SOFTWARE SOLUTIONS

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

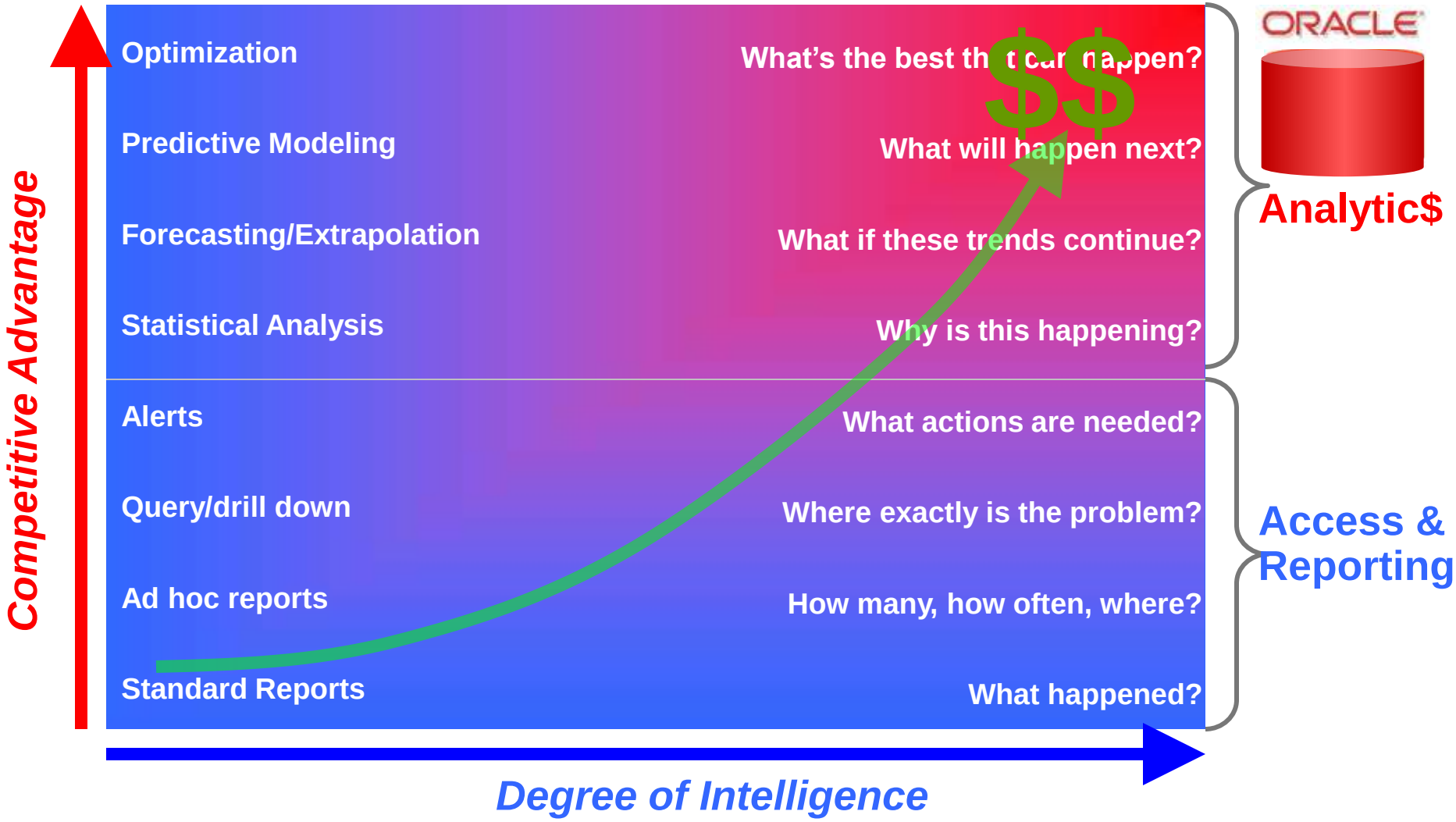
- Oracle OLAP
  - Defines a multi-dimensional data structure that allows information for highly complex calculations to done quickly.
  - Fast query performance and incremental update
  - Simplified access to analytic calculations
- Oracle Data Mining
  - Refers to the process of automatically sifting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

Spectrum of Oracle DB BI & Analytics

OLAP	Data Mining	Spatial
Summaries, trends and forecasts	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

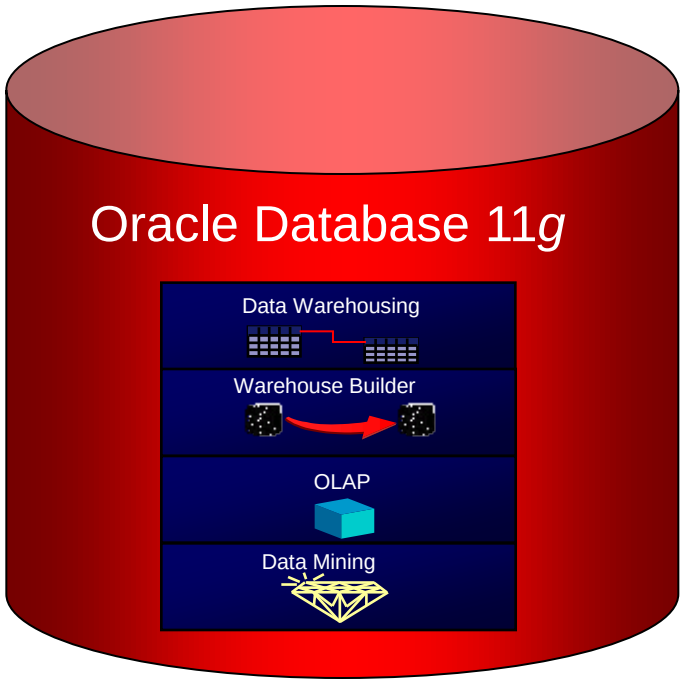


Source: Competing on Analytics, by T. Davenport & J. Harris

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

Oracle OLAP
Leveraging Core Database Infrastructure

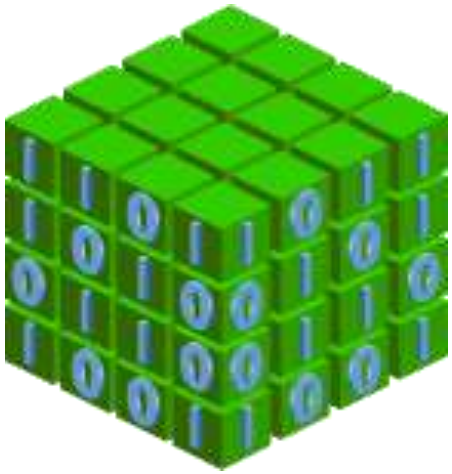


- Single RDBMS-MDBMS process
- Single data storage
- Single security model
- Single administration facility
- Grid-enabled
- Accessible by any SQL-based tool
- Embedded BI metadata
- Connects to all related Oracle data

vlamis SOFTWARE SOLUTIONS

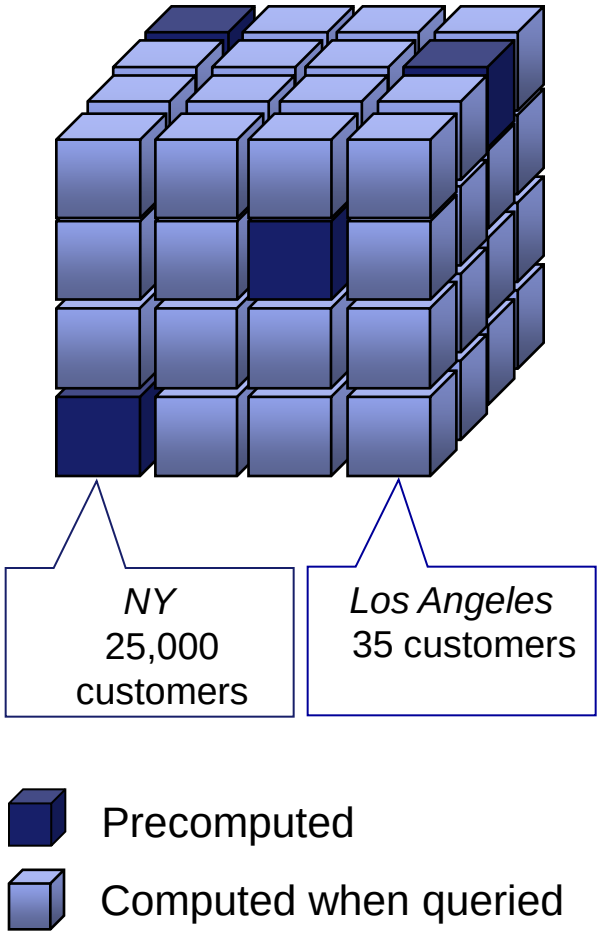
Oracle OLAP

- 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

Cost Based Aggregation
Pinpoint Summary Management



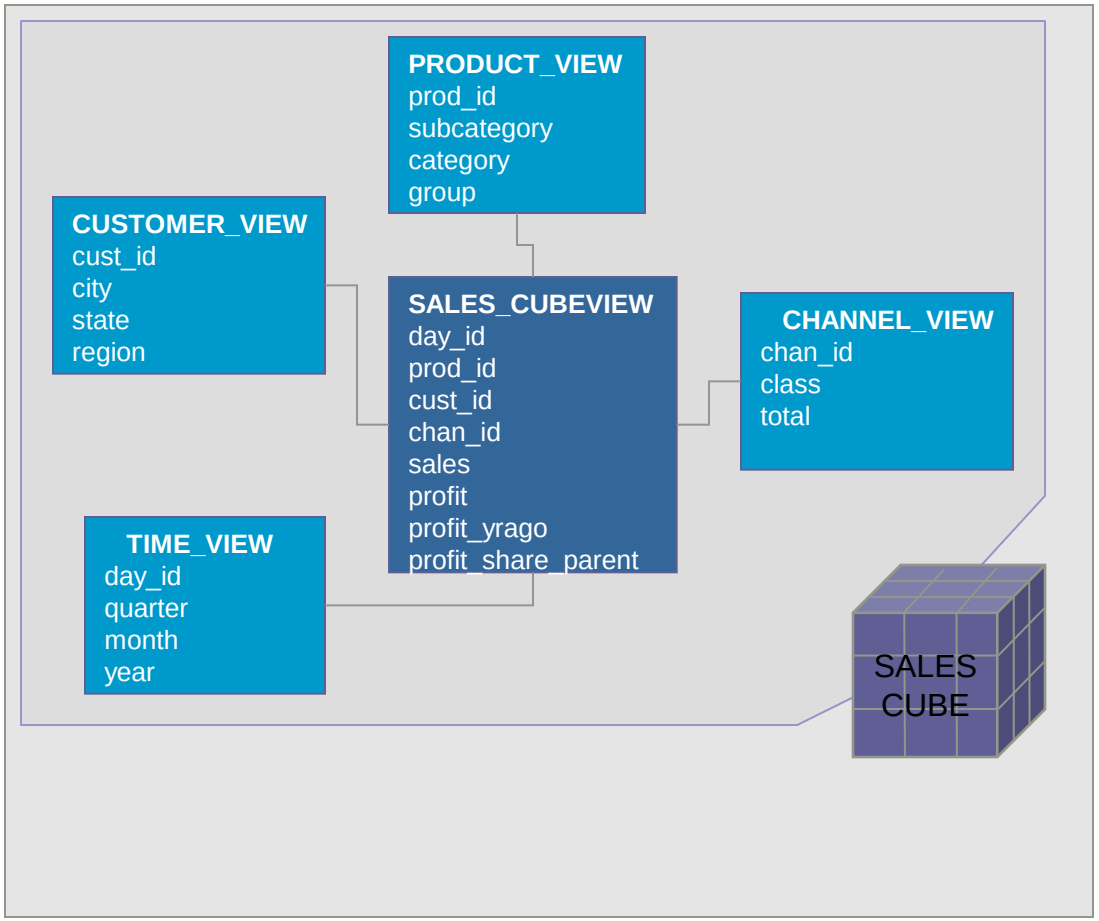
- Improves aggregation speed and storage consumption by pre-computing cells that are most expensive to calculate
- Easy to administer
- Simplifies SQL queries by presenting data as fully calculated

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

Cube Represented as Star Model
Simplifies Access to Analytic Calculations



- 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

In Database Data Mining

Traditional Analytics

- Data Import
- Data Mining Model "Scoring"
- Data Preparation and Transformation
- Data Mining Model Building
- Data Prep & Transformation
- Data Extraction

Oracle Data Mining

Savings

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

Results

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

Model "Scoring"
Data remains in the Database

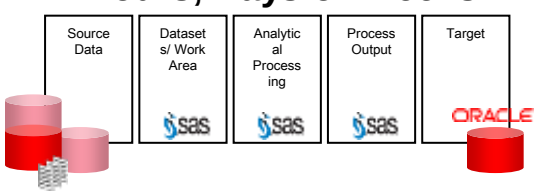
Embedded data preparation

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

Hours, Days or Weeks

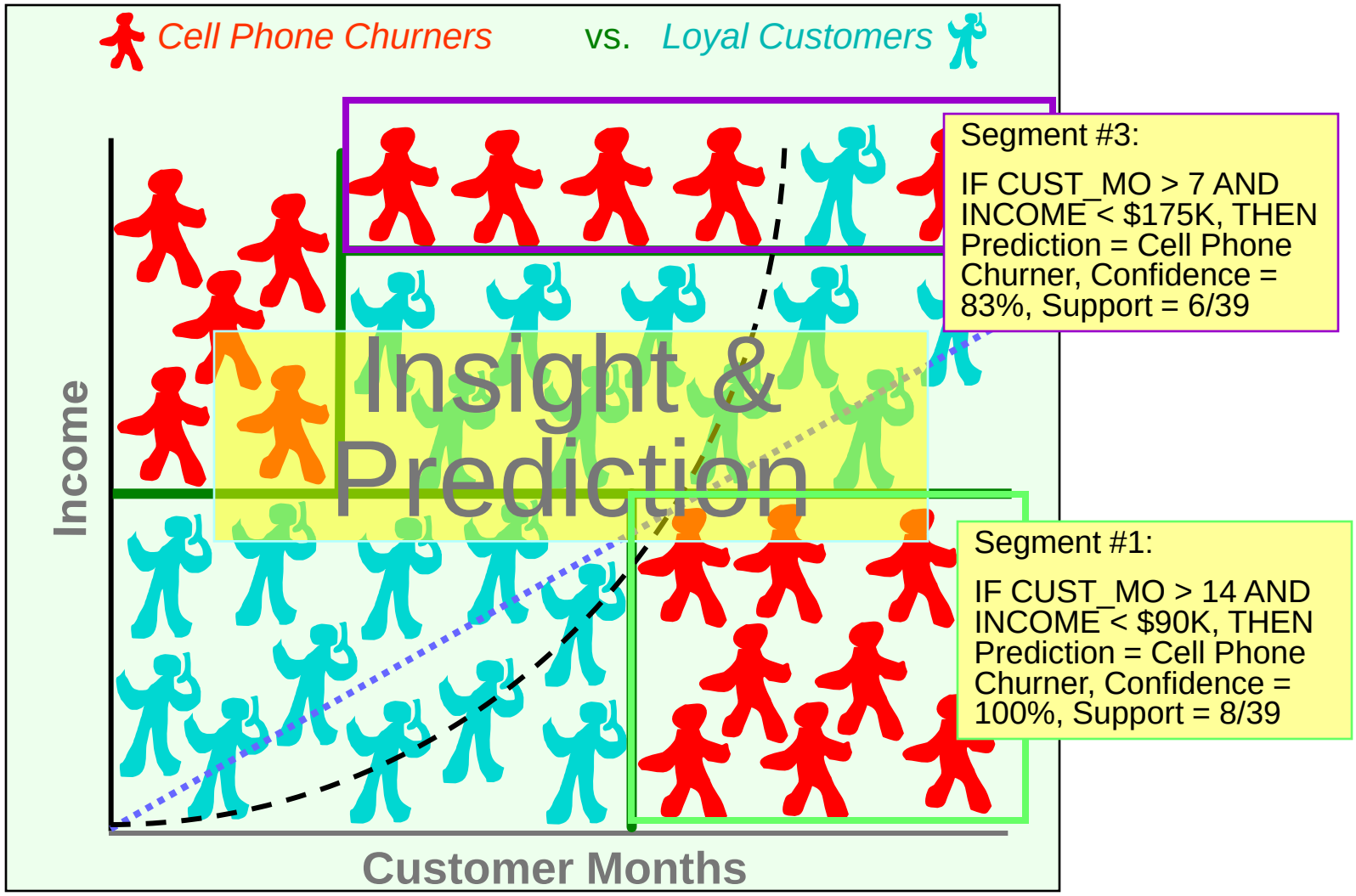


Secs. Mins or Hours

vlamis SOFTWARE SOLUTIONS

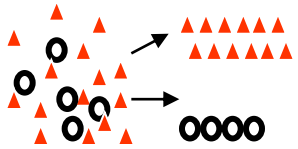
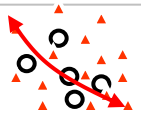
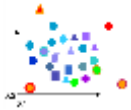
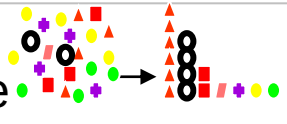
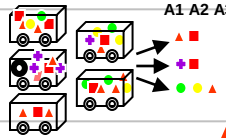
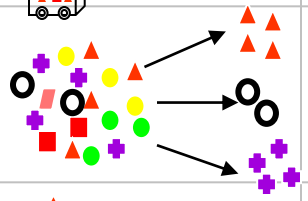
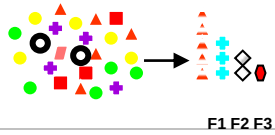
Data Mining Provides

Better Information, Valuable Insights and Predictions

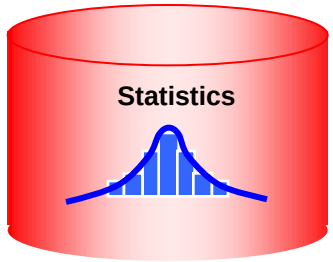


Source: Inspired from *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management* by Michael J. A. Berry, Gordon S. Linoff

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 of target field
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

11g Statistics & SQL Analytics (Free)



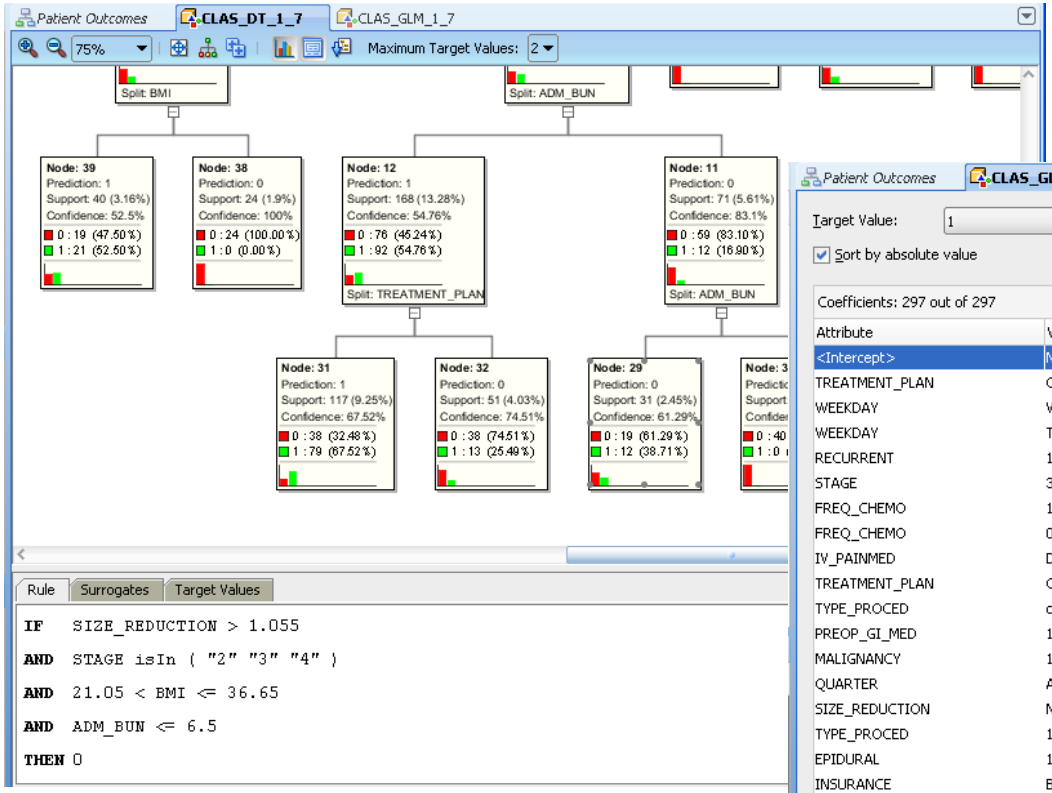
- Ranking functions
  - rank, dense_rank, cume_dist, percent_rank, ntile
- Window Aggregate functions (moving and cumulative)
  - Avg, sum, min, max, count, variance, stddev, first_value, last_value
- LAG/LEAD functions
  - Direct inter-row reference using offsets
- Correlations
  - Pearson's correlation 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
- 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

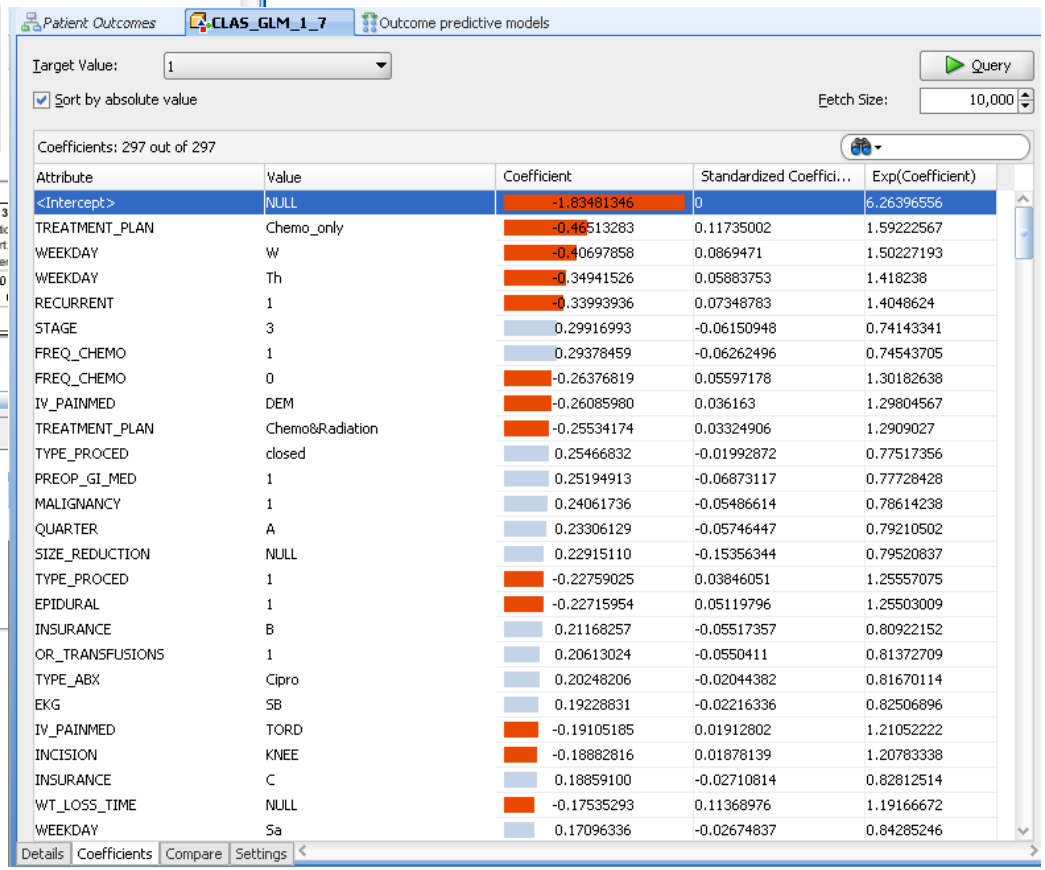
Descriptive Statistics

- DBMS_STAT_FUNCS: summarizes numerical columns of a table and returns count, min, max, range, mean, median, variance, standard deviation, quantile values, +/- n sigma values, top/bottom 5 values
- Reporting Aggregate functions
  - Sum, avg, min, max, variance, stddev, count, ratio_to_report
- 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
- Statistical Aggregates
  - Correlation, linear regression family, covariance
- Spatial Calculations
  - Distance, distance between, length, area

Note: Statistics and SQL Analytics are included in Oracle Database Standard Edition

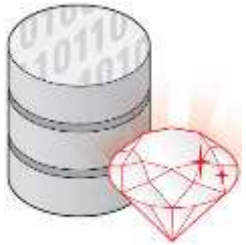
Understand Model Details





- Interactive model viewers

In-Database Data Mining



Traditional Analytics

- Data Import
- Data Mining Model "Scoring"
- Data Preparation and Transformation
- Data Mining Model Building
- Data Prep & Transformation
- Data Extraction

Oracle Data Mining

Savings

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

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

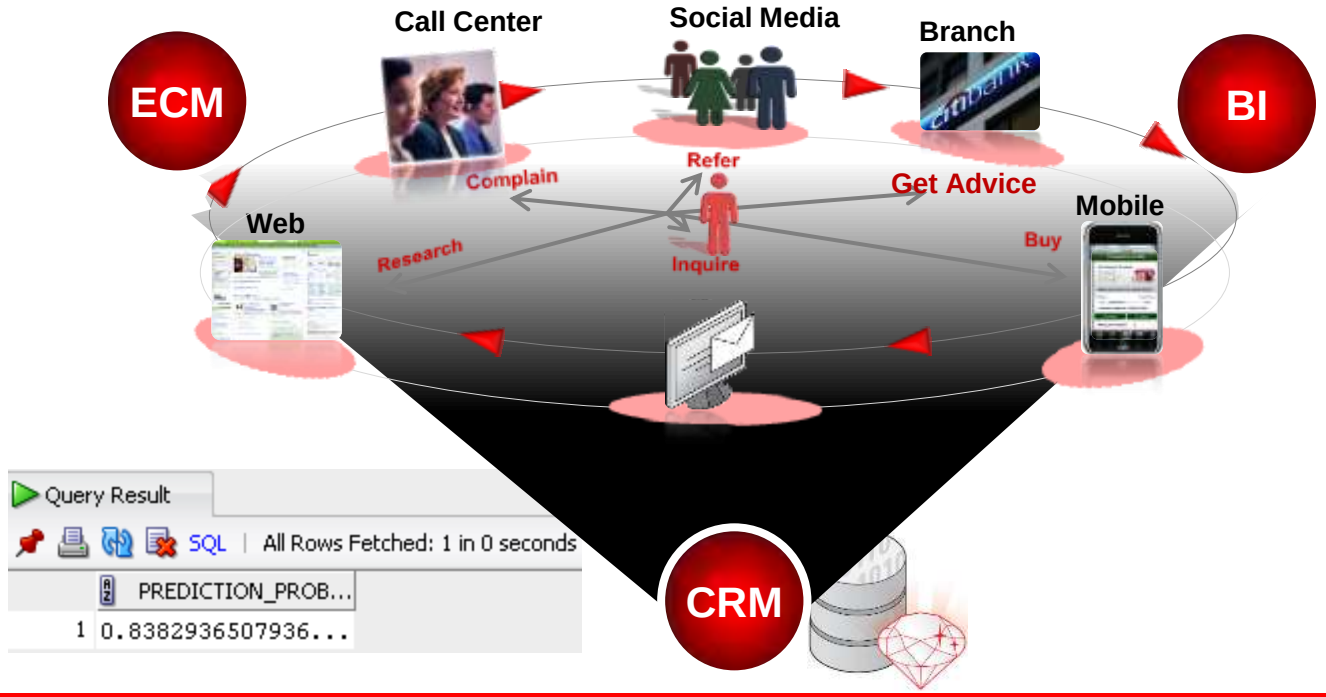
Hours, Days or Weeks

Secs. Mins or Hours

Real-time Prediction for a Customer

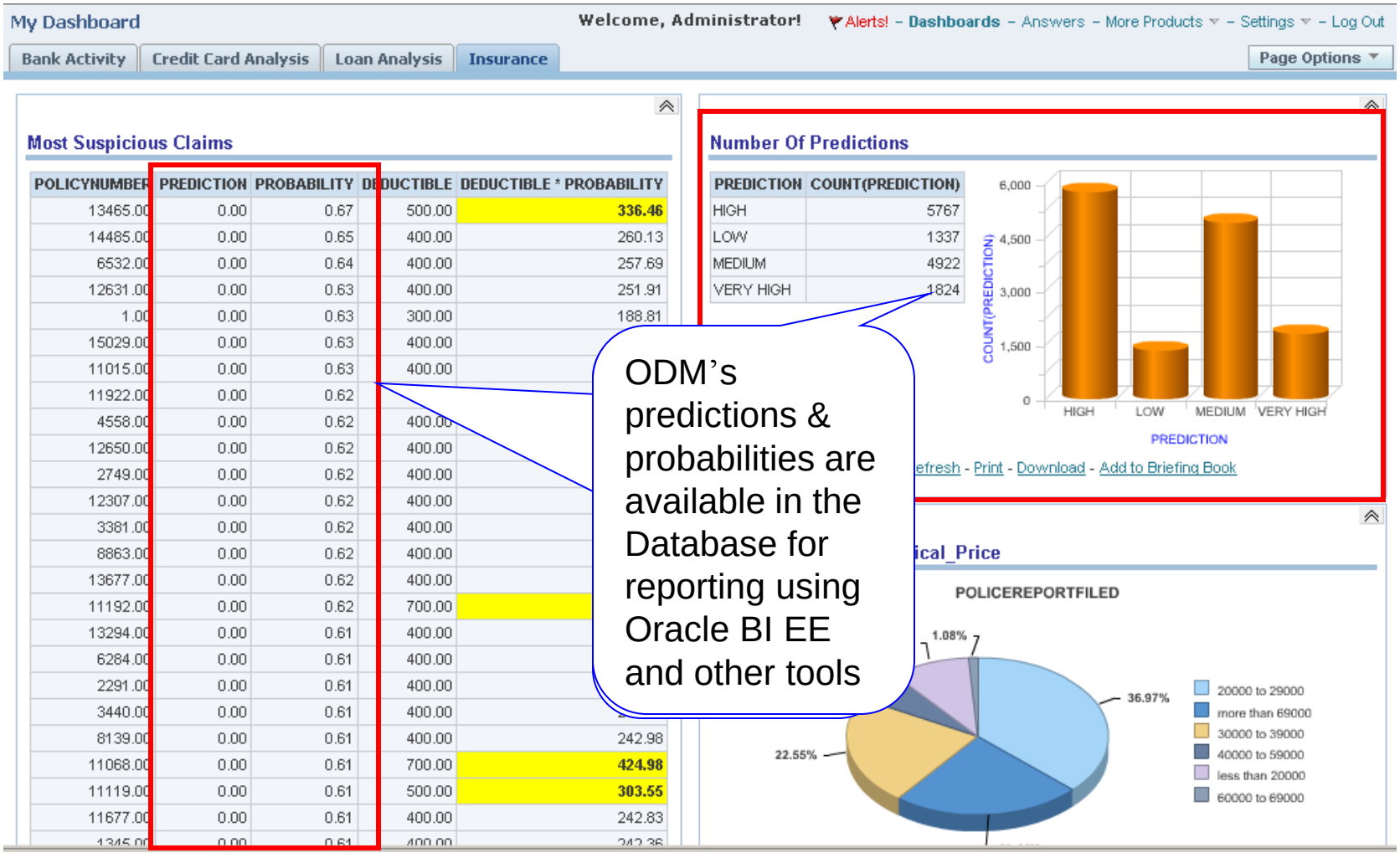
- On-the-fly, single record apply with new data (e.g. from call center)

```
Select prediction_probability(CLAS_DT_5_2, 'Yes'
  USING 7800 as bank_funds, 125 as checking_amount, 20 as
  credit_balance, 55 as age, 'Married' as marital_status,
    250 as MONEY_MONTLY_OVERDRAWN, 1 as house_ownership)
from dual;
```



Example

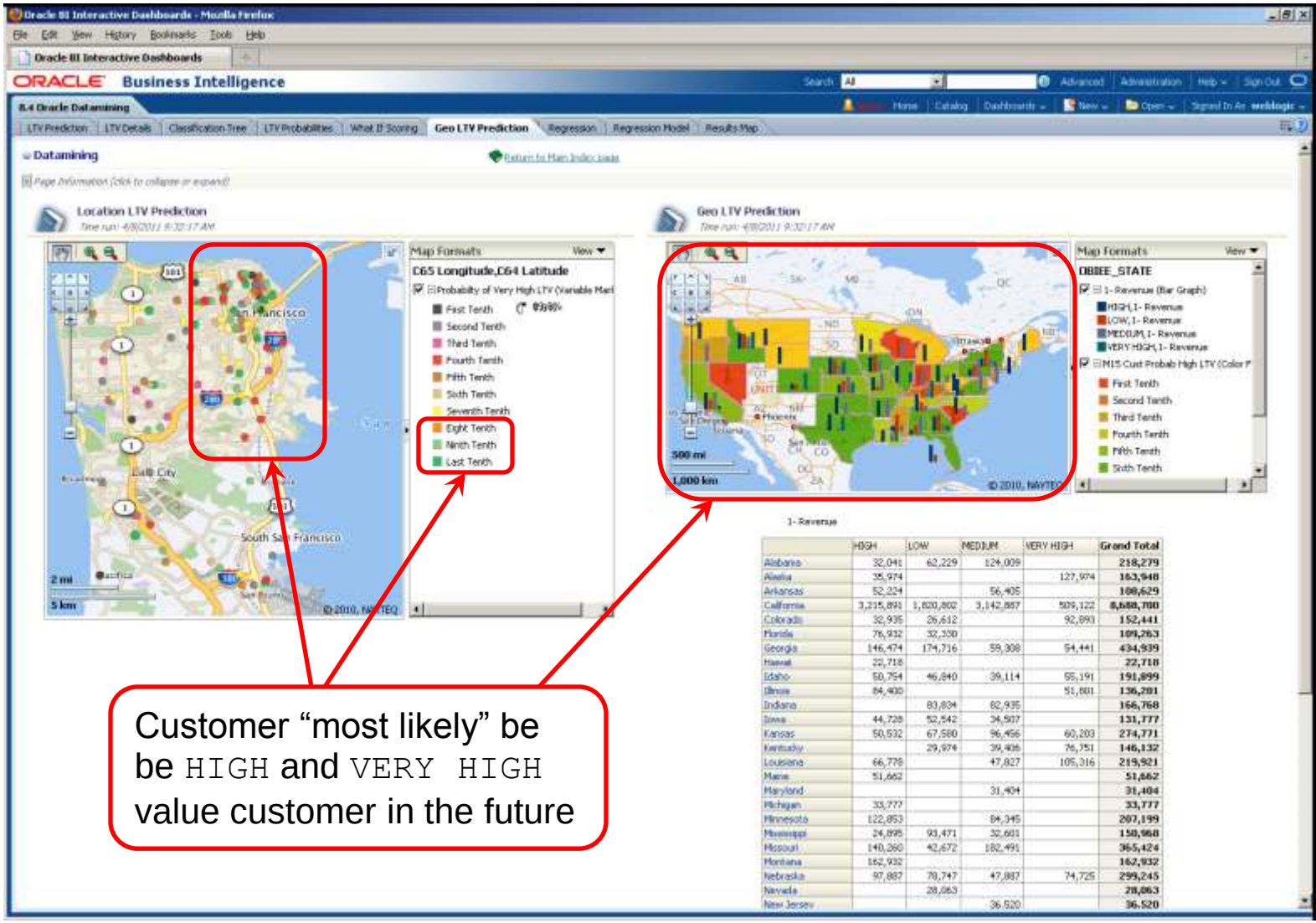
Better Information for OBI EE Reports and Dashboards



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

Combination of Data Mining and Spatial

- In-database data mining builds predictive models that predict customer behavior
- OBIEE's integrated spatial mapping shows where



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.

Natively Manage All Geospatial Data

"Points"

"Lines"

"Polygons"

Web Services (OGC)

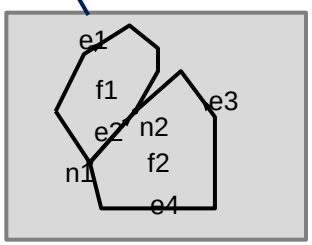
Oracle database

ORACLE SPATIAL

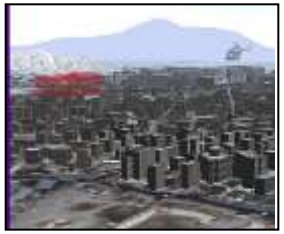
Geocoding Routing

Rasters

Networks



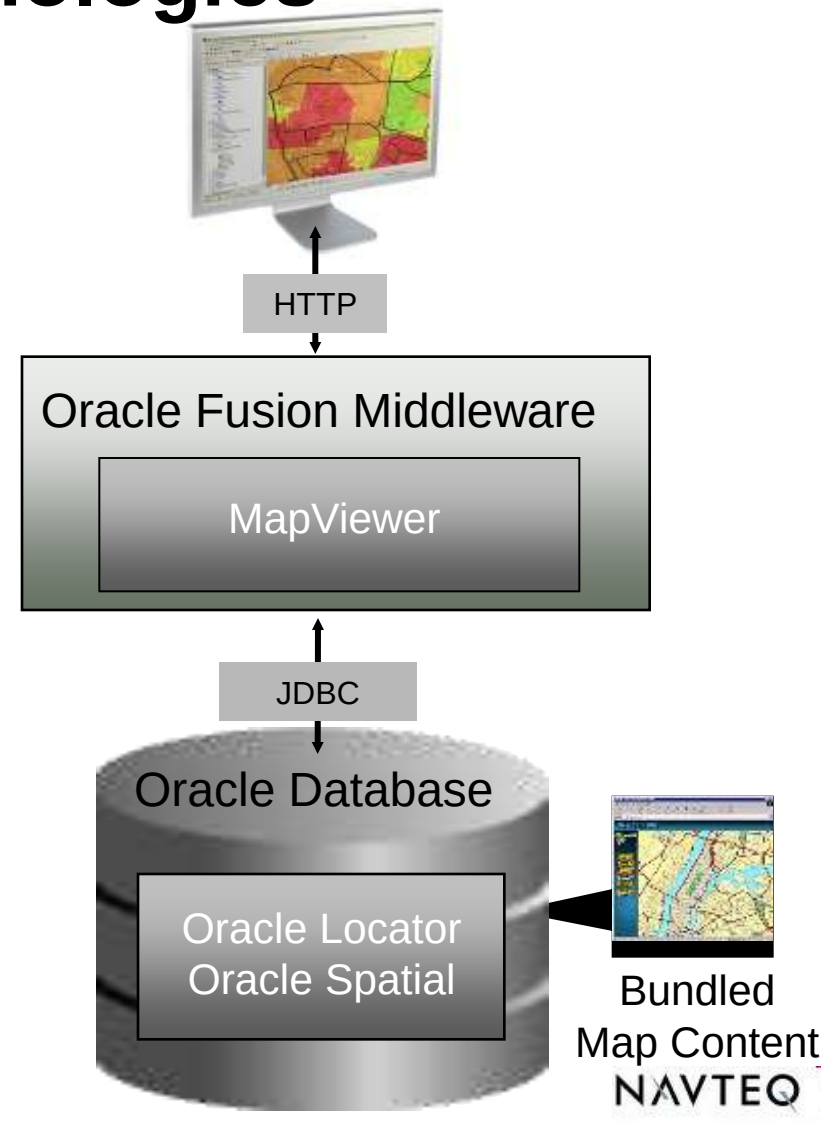
Topologies



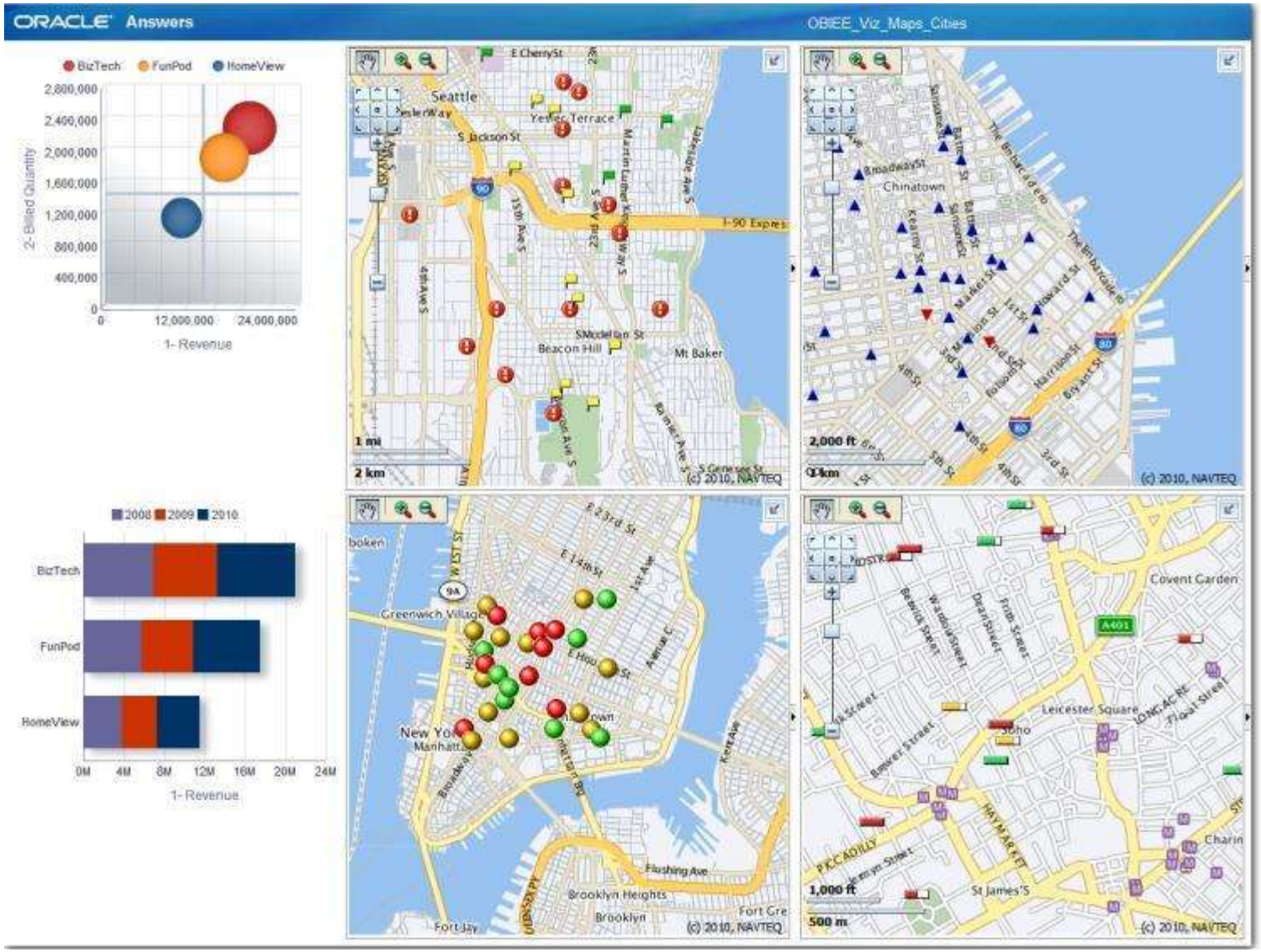
3D

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



Depict and Detect Spatial Relationships

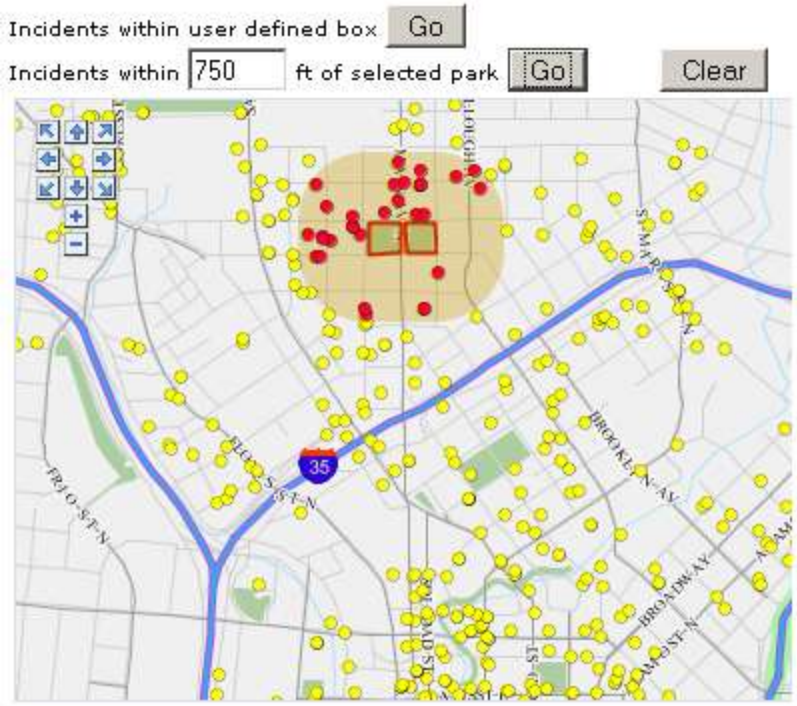


Why Maps are Powerful

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

vlamis SOFTWARE SOLUTIONS

Some Analysis Is Possible Only with Spatial Analytics



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

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.

Oracle Fusion Middleware MapViewer

- A J2EE component (.ear) for developing web mapping applications. Usually deployed in WLS.
- Renders geospatial content stored in an Oracle database. Background maps can be from 3rd party providers.
- Provides Java, Javascript, and XML request/response APIs.