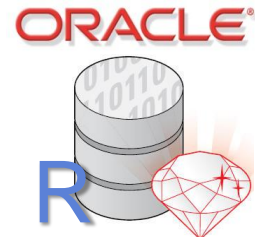


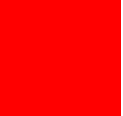


ORACLE®

Oracle Advanced Analytics

Oracle R Enterprise & Oracle Data Mining





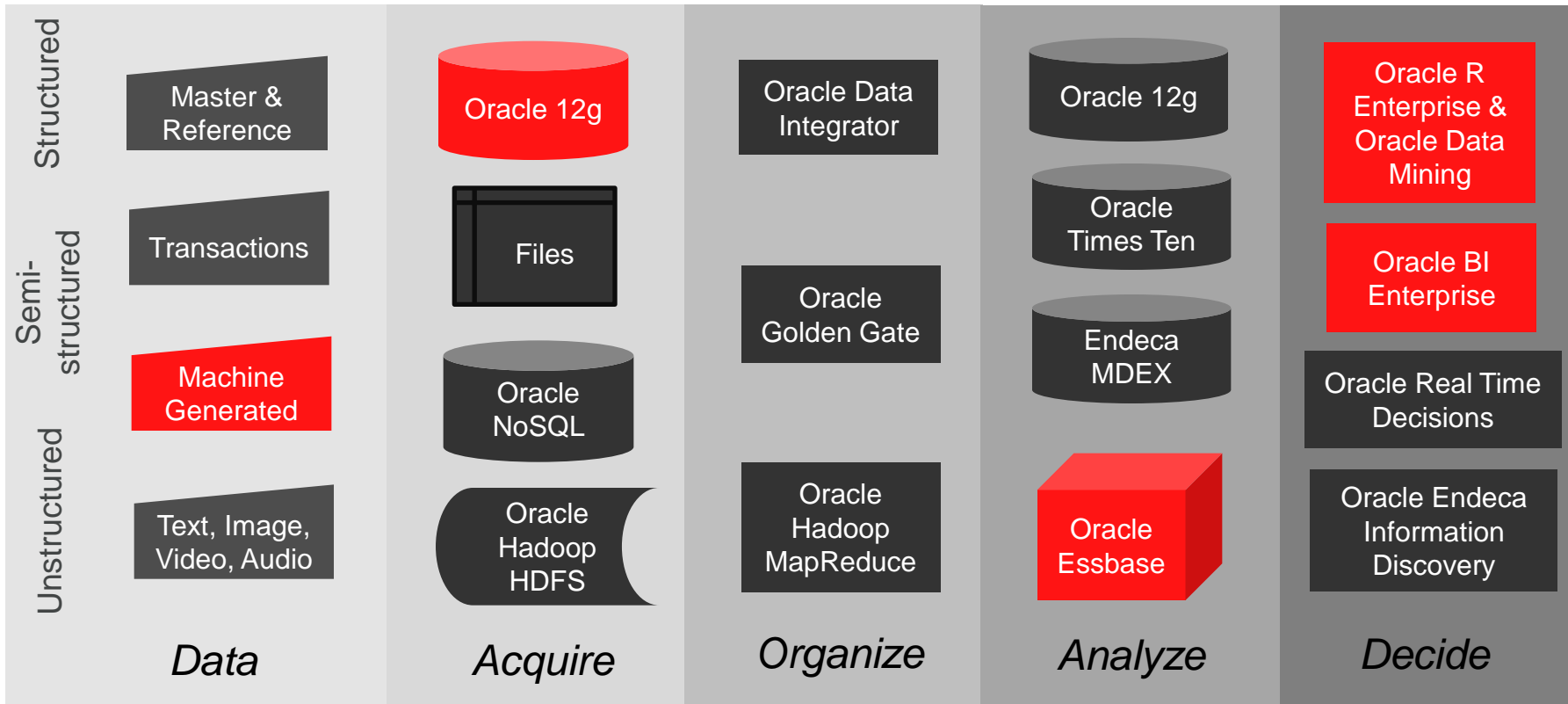
The following is intended to outline our general product direction. It is intended for information purposes only, and may not be incorporated into any contract. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions.

The development, release, and timing of any features or functionality described for Oracle's products remains at the sole discretion of Oracle.

USE CASE 8: ELECTRICAL CONSUMPTION

- Short Description :
 - CERN ENS (Electrical Network Supervisor) system archives values from measurements
 - Around 12000 every 10 minutes
 - The storage is managed by an external company on an instance of Oracle relational database.
 - Nevertheless, the DB instance is installed on one dedicated server connected to CERN Technical Network
- Issues:
 - Make data aggregation (by time and by families of measurements)
 - Electricity load forecast by system (families of measurements) using the historical data
 - Make the result obtained accessible to CERN users from General Purpose Network

Oracle Technology mapped to Analytics Landscape



Oracle Advanced Analytics Option—Agenda

Extending the Database into a Comprehensive Advanced Analytics Platform

- Oracle Data Mining
 - SQL & PL/SQL focused in-database data mining and predictive analytics
- Oracle R Enterprise
 - Integrates Open Source R with the Oracle Database



ORACLE

Oracle Advanced Analytics Option

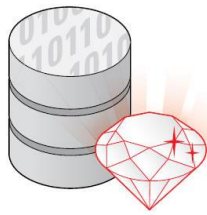
Transforming the Database into a Comprehensive Advanced Analytics Platform



- Oracle Advanced Analytics Option enables companies to *"bring the algorithms to the data"* vs. extracting the data to specialized and expensive dedicated statistical and data mining servers
- Oracle Advanced Analytics Option includes:
 - Oracle Data Mining
 - SQL & PL/SQL focused in-database data mining and predictive analytics
 - Oracle R Enterprise
 - Integrates the Open-Source Statistical Environment R with the Oracle Database
- Data movement is eliminated or dramatically reduced while analytical and compute intensive operations are performed inside the database

Oracle Data Mining

Building Predictive Analytics Applications

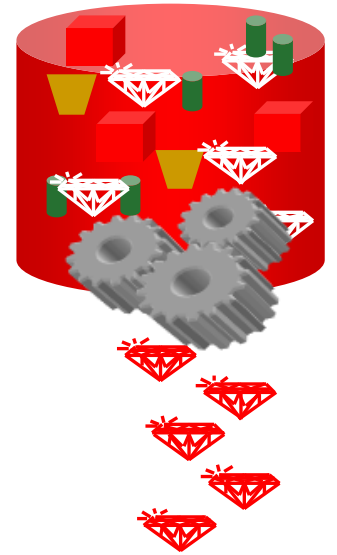


- Oracle Data Mining provides 12 powerful in-database data mining algorithms for big data analytics as a native feature of the database
 - Designed for or big data problems involving discovering patterns and relationships in large amounts of data and oftentimes making predictions based on those patterns, Oracle Data Mining allows data analysts and data miners to mine star schemas, transactional data and unstructured data stored inside the database, build predictive models and apply them to data inside the database--all without moving data.
- Developers can use the Oracle Data Miner extension to SQL Developer to develop, build, evaluate, share and automate analytical workflows to solve important data driven business problems.
- Developers can use the SQL APIs and PL/SQL to build applications to automate knowledge discovery
 - The Oracle Data Miner GUI generates SQL code that application developers can use to develop and deploy SQL and PL/SQL based automated predictive analytics applications that run natively inside the Oracle Database.

What is Data Mining?

- Automatically finds 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*)

ORACLE®



ORACLE®

SQL Developer 3.0/Oracle Data Miner 11g Release 2 GUI

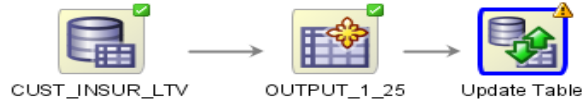
- Graphical User Interface for data analyst
- SQL Developer Extension (OTN download)
- Explore data—discover new insights
- Build and evaluate data mining models
- Apply predictive models
- Share analytical workflows
- Deploy SQL Apply code/scripts

The screenshot displays the Oracle SQL Developer 3.0 GUI with the Oracle Data Miner extension. The main workspace shows a workflow diagram with nodes: 'CUST_INSUR_LTV', 'Explore Data 2', 'Column filter and AI', 'Clust Build 21', '5 Response Models', 'Model Details 17', and 'Apply 14'. A red starburst graphic with the text 'New GUI' is overlaid on the workflow. The 'Model Details 17' window is open, showing a table of model settings.

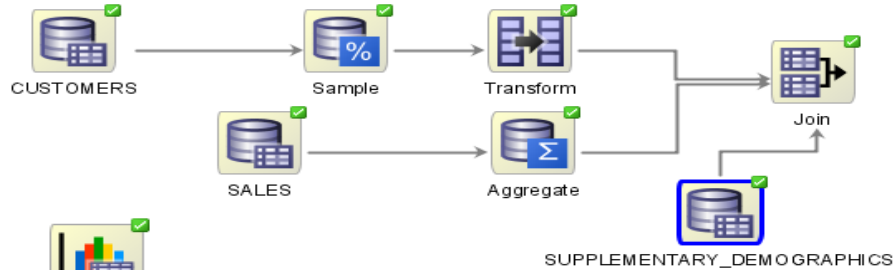
Name	Build	Test	Tune	Algorithm	Comment
CLAS_GLM_...	7/13/10 6:07...	7/13/10 6:07...	Automatic	Generalized Line...	
CLAS_SVM_...	7/13/10 6:06...	7/13/10 6:06...	Automatic	Support Vector ...	
CLAS_SVM_...	7/13/10 6:06...	7/13/10 6:07...	Automatic	Support Vector ...	
CLAS_DT_3_3	7/13/10 6:06...	7/13/10 6:06...	Automatic	Decision Tree	
CLAS_NB_3_3	7/13/10 6:06...	7/13/10 6:06...	Automatic	Naive Bayes	

Oracle Data Miner Nodes (Partial List)

Tables and Views



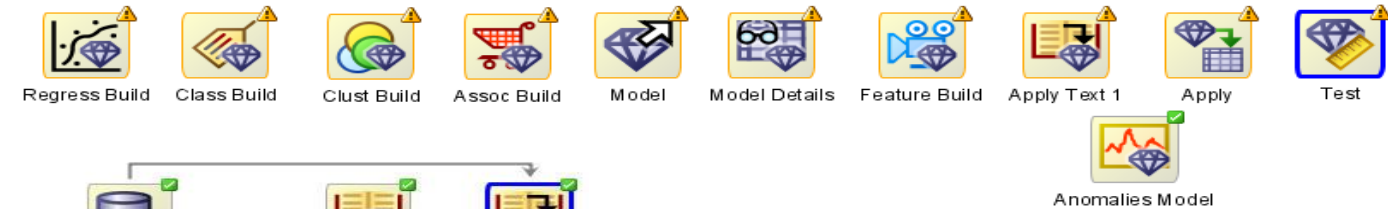
Transformations



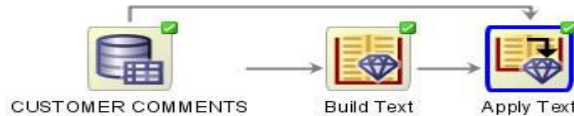
Explore Data



Modeling

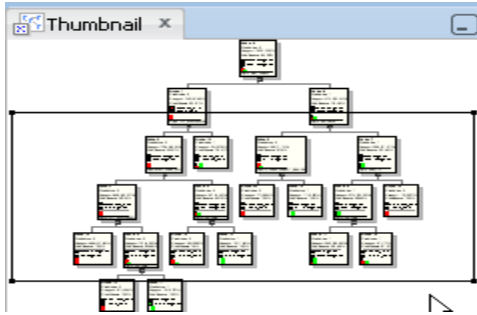


Text

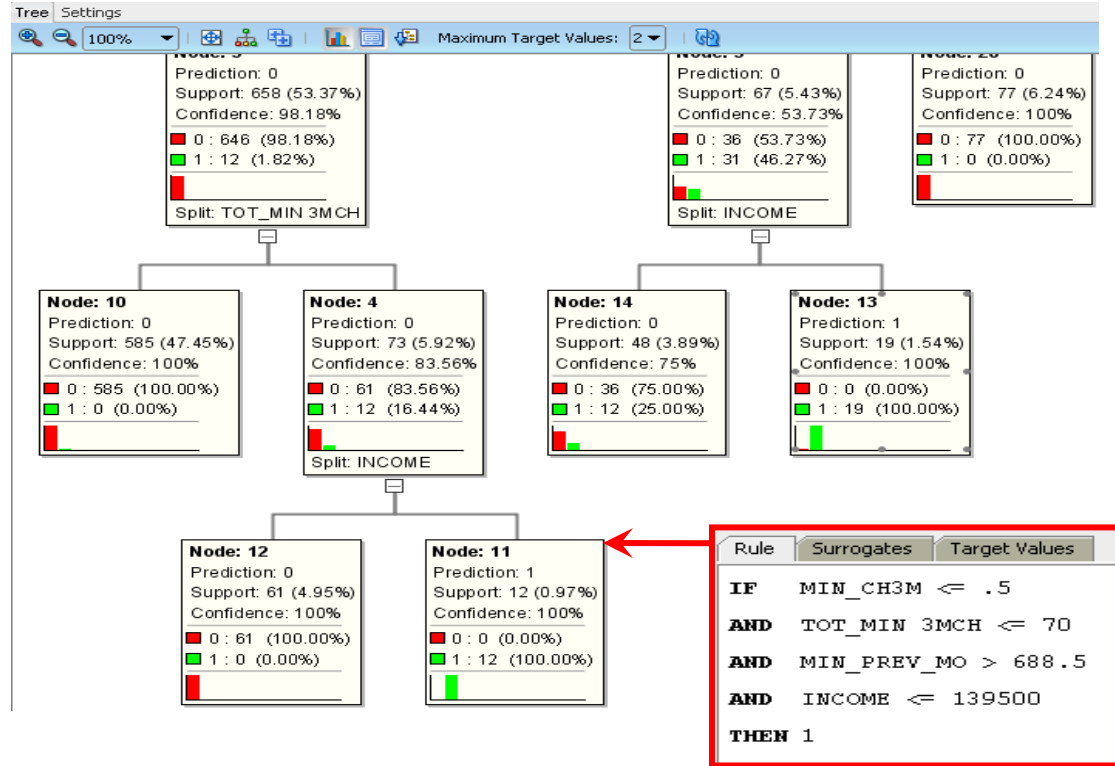


Oracle Data Miner 11g Release 2 GUI

Churn Demo—Simple Conceptual Workflow

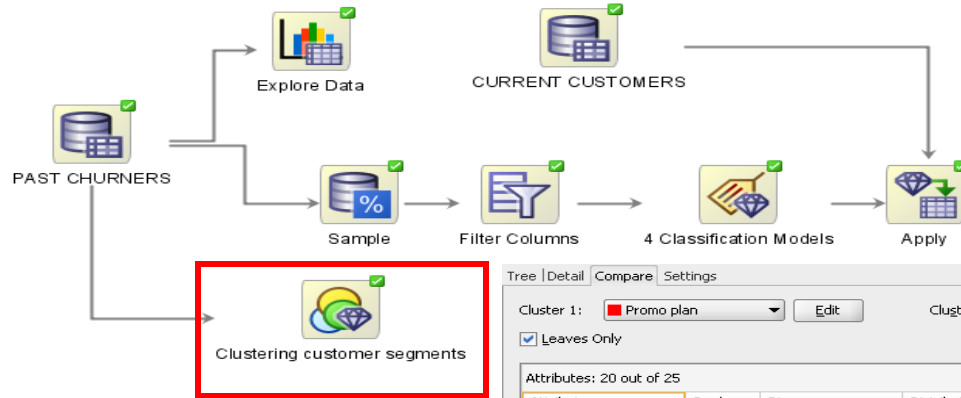


Churn models to product
and “profile” likely
churners



Oracle Data Miner 11g Release 2 GUI

Simple Conceptual Workflow



Clustering analysis to discover customer segments based on behavior, demographics, plans, equipment, etc.

Cluster 1: ■ Promo plan Edit Cluster 2: ■ Local_domestic Edit

Leaves Only Fetch Size: []

Attributes: 20 out of 25 Attrib

Attribute	Rank	Divergence	Distribution	Centroid(12)	Centroid(15)
CONVERGENT_BILLING	12	0.4445		No	Yes
CHURNER	13	0.3463		0	0.70588235
MIN_CH1M	14	0.1147		394.44444444	-95.52941176
MIN_CH3M	15	0.0455		-510.88888889	970.82352941
MIN_CURR_MO	16	0.0426		489.44444444	1,913.88235294
TOT_MIN_CHNG	16	0.0426		-116.44444444	875.29411765

PHONE_PLAN

Cluster Promo plan (red) Cluster Local_domestic (blue)

Fraud Prediction Demo

```
drop table CLAIMS_SET;  
exec dbms_data_mining.drop_model('CLAIMSMODEL');  
create table CLAIMS_SET (setting_name varchar2(30), setting_value varchar2(4000));  
insert into CLAIMS_SET values  
    ('ALGO_NAME','ALGO_SUPPORT_VECTOR_MACHINES');  
insert into CLAIMS_SET values ('PREP_AUTO','ON');  
commit;
```

```
begin  
dbms_data_mining.create_model('CLAIMSMODEL', 'CLASSIFICATION',  
    'CLAIMS', 'POLICYNUMBER', null, 'CLAIMS_SET');  
end;  
/
```

```
-- Top 5 most suspicious fraud policy holder claims  
select * from  
(select POLICYNUMBER, round(prob_fraud*100,2) percent_fraud,  
    rank() over (order by prob_fraud desc) rnk from  
(select POLICYNUMBER, prediction_probability(CLAIMSMODEL, '0' using *) prob_fraud  
from CLAIMS  
where PASTNUMBEROFCLAIMS in ('2to4', 'morethan4'))  
where rnk <= 5  
order by percent_fraud desc;
```

POLICYNUMBER	PERCENT_FRAUD	RNK
6532	64.78	1
2749	64.17	2
3440	63.22	3
654	63.1	4
12650	62.36	5

Automated Monthly “Application”! *Just add:*

```
Create  
View CLAIMS2_30  
As  
Select * from CLAIMS2  
Where mydate > SYSDATE – 30
```



ORACLE

Exadata + Data Mining 11g Release 2

“DM Scoring” Pushed to Storage!



- In 11g Release 2, SQL predicates and Oracle Data Mining models are pushed to storage level for execution

For example, find the US customers likely to churn:

```
select cust_id
from customers
where region = 'US'
and prediction_probability(churnmod, 'Y' using *) > 0.8;
```

Scoring function executed in Exadata

Oracle Communications Industry Data Model

Better Information for OBIEE Dashboards

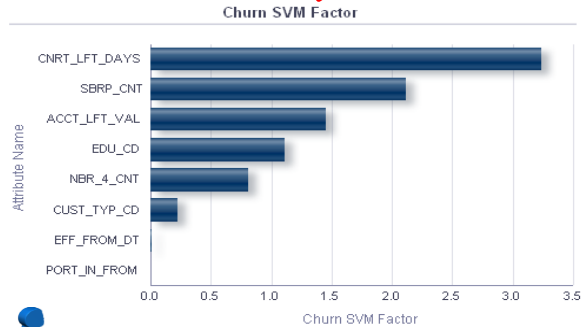
ORACLE Business

Customer Churn Analysis

Oracle Data Mining identifies key contributors to customer churn

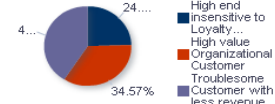
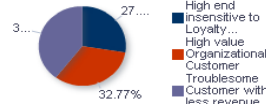
What's the revenue loss of the churned customer?
Click Here...

Churn Factor Rank



Attribute Name	Attribute Value	Rank	Churn SVM Factor
CNRT_LFT_DAYS		1	3.2345
SBRP_CNT		2	2.1112
ACCT_LFT_VAL		3	1.4523
EDU_CD	HIGH_SCHOOL	4	1.1112
NBR_4_CNT		5	0.8112
CUST_TYP_CD		6	0.2251
EFF_FROM_DT		7	0.1112
PORT_IN_FROM		8	0.1112

Churn Profile



Oracle Data Mining discovers different profiles of churning customers and their profile, both critical to developing a proactive response to reduce churn.

Customer Segment	Customer Segment	ARPU Band	Customer Count	Churned Customer Count	Actual Revenue last Month	Avg Debt Value
High end insensitive to Loyalty Program	CUST_TYP_CD is IND; LYLTY_PROG_BAL=773.81; AGE_ON_NET_NBR=1975.87; MO_RVN=406;	ARPU7500 BAND	3	0	\$7,000	\$130
		ARPU5000 BAND	24	0	\$74,656	\$3,734
		ARPU2500 BAND	3	0	\$6,588	\$422
		ARPU1000 BAND	443	54	\$132,363	\$69,845
High value Organizational Customer	CUST_TYP_CD is ORG; SBRP_CNT=85.3; AGE_ON_NET_NBR=923.72; TOT_RVN=39,942;	ARPU7500 BAND	4	1	\$25,764	\$659
		ARPU5000 BAND	35	3	\$96,908	\$5,059
		ARPU2500 BAND	9	1	\$21,314	\$1,287
		ARPU1000 BAND	519	54	\$166,594	\$80,870

ORACLE

Oracle Communications Industry Data Model Example

Better Information for OBIEE Dashboards

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

ORACLE Business Intelligence

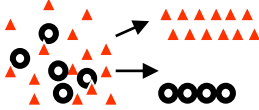


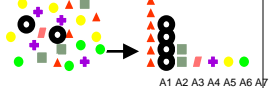
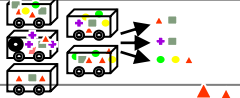
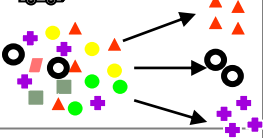
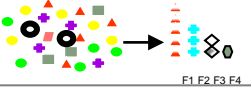
Churn Report By customer Segment

Customer Segments Customer Segmentation Details

Customer Segment Name is equal to Age Young and PAY TV


Customer Segment	Customer Name	Cell Phone No	Contract Value	Month Revenue	Debt Value	LTV Band	LTV Value	LTV Months	ARPU Band	Churn Indicator	Sentiment	Churn Probability	Customer Segment Key	
Age Young and PAY TV user	Beverly Wan	9985007046	\$9,000.00	\$7,800.00	\$70.00	LTV_1	\$41,000.00	44	ARPU7500+		▲ +	59	104	
	Bradley Johnson	9985007589	\$9,000.00	\$8,100.00	\$222.00		\$49,000.00	32			▲ +	45	104	
	Ethan Nielley	9985006289	\$9,000.00	\$8,400.00	\$70.00		\$34,000.00	43		● Probability of Churning is very high	▼ -	71	104	
	Tobias Hamrick	9985008239	\$0.00	\$7,800.00	\$130.00		\$69,000.00	9			▲ +	43	104	
	Gale Lazar	9985003794	\$9,000.00	\$7,000.00	\$70.00		\$82,000.00	37	ARPU7500		▲ +	16	104	
	Mallory Lawson	9985008346	\$0.00	\$5,400.00	\$130.00		\$63,000.00	30			▲ +	57	104	
	Abbie Anderson	9985010557		\$0.00	\$2,769.23	\$222.00		\$90,000.00	14	ARPU5000	● Probability of Churning is very high	▼ -	79	104
				\$0.00	\$2,769.23	\$222.00		\$99,000.00	18		● Probability of Churning is very high	▼ -	85	104

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 	Nonnegative Matrix Factorization	Text analysis Feature reduction

Learn More

Oracle Data Mining on OTN



Welcome Charles | Account | Help | Sign Out | United States | Communities | I am a... | I want to...

Products and Services | Downloads | Store | Support | Education | Partners | About

Oracle Technology Network > Database > Database Options > Data Mining

- Database 11g
- Database Focus Areas
- Database Options
- Database Express Edition
- Database Upgrade
- Berkeley DB
- TimesTen In-Memory Database
- Database 10g
- Database Lite
- RDB
- Audit Vault
- Secure Backup
- Real Application Clusters
- Clusterware
- Multimedia
- Windows

Oracle Data Mining

Powering Next-Generation Predictive Applications


[Technical & General Info](#), [Learn More](#), [Resources](#), [Related Technologies](#)

Oracle Data Mining (ODM)—a priced option to Oracle Database 11g Enterprise Edition—enables you to easily build and deploy next-generation applications that deliver predictive analytics and new insights. Because data models and results remain in the Oracle Database, data movement is eliminated, security is maximized and information latency is minimized. Oracle Data Mining models can be included in SQL queries and embedded in applications to offer improved business intelligence.

Application developers can rapidly build next-generation applications using ODM's SQL and Java APIs that automatically mine Oracle data and deploy results in real-time-throughout the enterprise. Because the data, models and results remain in the Oracle Database, data movement is eliminated, security is maximized and information latency is minimized. Oracle Data Mining models can be included in SQL queries and embedded in applications to offer improved business intelligence.

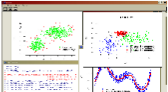
Oracle Data Mining provides in-Database predictive analytics that support strategies described in the Harvard Business Review (HBR) article [Competing on Analytics](#). Data analysts can quickly access their Oracle data using the optional [Oracle Data Miner](#) graphical user interface and explore their data to find patterns, relationships, and hidden insights. Oracle Data Mining provides a collection of in-database data mining algorithms that solve a wide range of business problems. Anyone who can access data stored in an Oracle Database can access Oracle Data Mining results—predictions, recommendations, and discoveries using SQL-based query and reporting tools including [Oracle Business Intelligence EE Plus](#).

New: [Oracle Data Mining at Oracle OpenWorld 2010](#)



New: [R Interface to Oracle Data Mining](#)

The R Interface to Oracle Data Mining (R-ODM) allows R users to access the power of Oracle Data Mining's in-database functions using the familiar R syntax. R-ODM provides a powerful environment for prototyping data analysis and data mining methodologies.



Oracle Data Mining Blog

BLOGS HOME | PRODUCTS & SERVICES | INDUSTRIES | SUPPORT | PARTNERS | COMMUNITIES | ABOUT

Oracle Data Mining (ODM)

Everything about Oracle Data Mining - News, Technical Information, Opinions, Tips & Tricks. All in One Place

To sample or not to sample... [Part 4](#) < [main](#) > [The Meaning of Probability](#)

Oracle Data Miner 11g Release 2 Update: Now Extension to SQL Developer

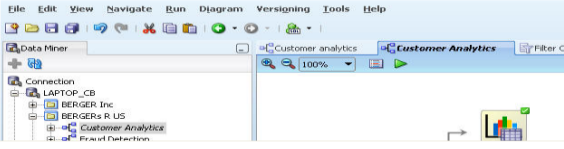
By [charlieberger](#) on July 15, 2010 8:43 PM

News: The Oracle Data Miner 11g Release 2 New "Work Flow" GUI is now being packaged as an Extension to SQL Developer and will be available to external customers as part of the SQL Dev. 3.0 next release Early Adopter program. SQL Developers will be able to access Oracle Data Miner's data mining GUI from within the familiar SQL Developer environments. This tight integration will provide a number of significant advantages for data analyst, developers and DBAs including:

- Everything - Data access, SQL querying and data transformations and Data Mining functionality all in a complete, unified environment - inside the Oracle Database
- Elimination of data movement, loss of security, and information latency to extract data to traditional external data analysis servers e.g. SAS, SPSS.
- Ability to create and deploy complex predictive analytics methodologies within the Oracle SQL Developer environment
- Ability to [Click for Updates](#) and get the latest version of Oracle Data Miner 11g Release 2 GUI
- Access to [Oracle By Examples](#) (OBE) posted on OTN

Stayed tuned to [Oracle SQL Developer on the Oracle Technology Network \(OTN\)](#) and [Oracle Data Mining on OTN](#) web site and this for updates and more information.

[Sample Oracle Data Miner 11g Release 2 New "Work Flow" GUI screen shots.](#)



Search All Blogs

Google Search

Only search this blog

About This Entry

This page contains a single entry from the blog posted on July 15, 2010 04:3 PM

The previous post in this blog was [To sample or not to sample... Part 4](#)

The next post in this blog is [The Meaning of Probability](#)

Many more can be found on the [main index page](#) or by looking through the [archives](#)

Top Tags

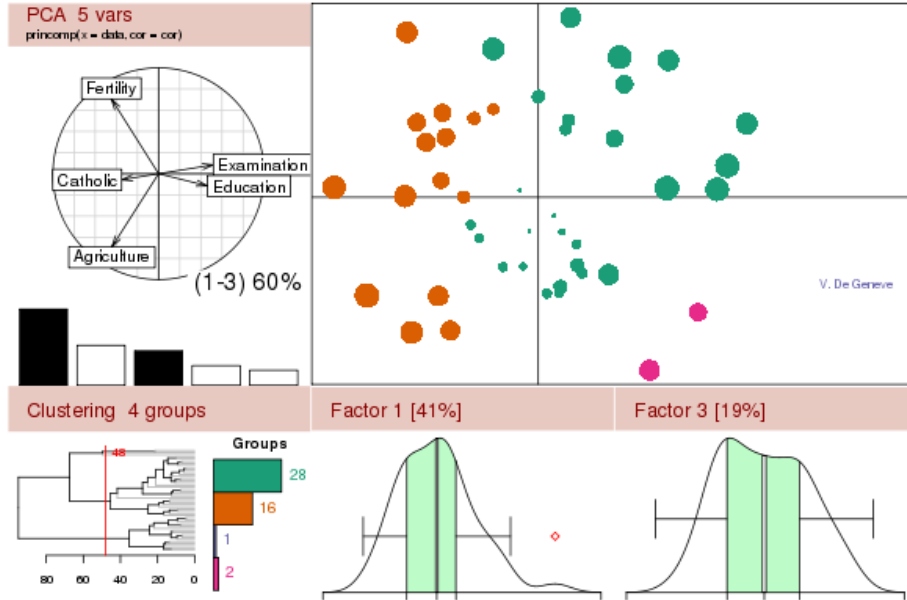
11g Release 2 actionable insight
Amazon EC2 America's Cup Analytics
anomaly detection application
applications extension AUVS BI BI/IVA
BI/VA BI/IV BI/IVM BI/IVM Cloud Computing
Clustering Data Mining
data mining data
transformation database DBA
Distributed Data Mining Environment
Clustering Database OracleData Forecaster
Fraud funny graphics industry data
model innovative holistic mining
value Model Deployment Model
Database Model Export Model Import
Model Transport raw GUI NPL OBI/EE
open sql ODM user home
OpenWorld Oracle Oracle Business
Intelligence EE Oracle Data
Miner oracle data miner 11gR2



Oracle Data Mining



R Statistical Programming Language



Open source language and environment

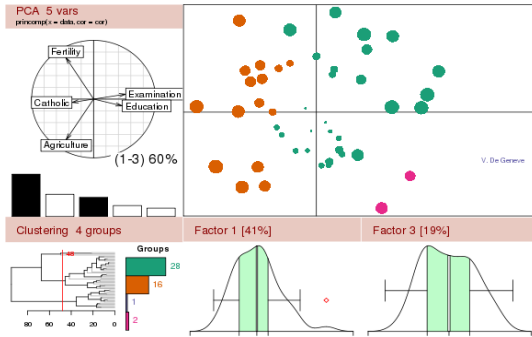
Used for statistical computing and graphics

Strength in easily producing publication-quality plots

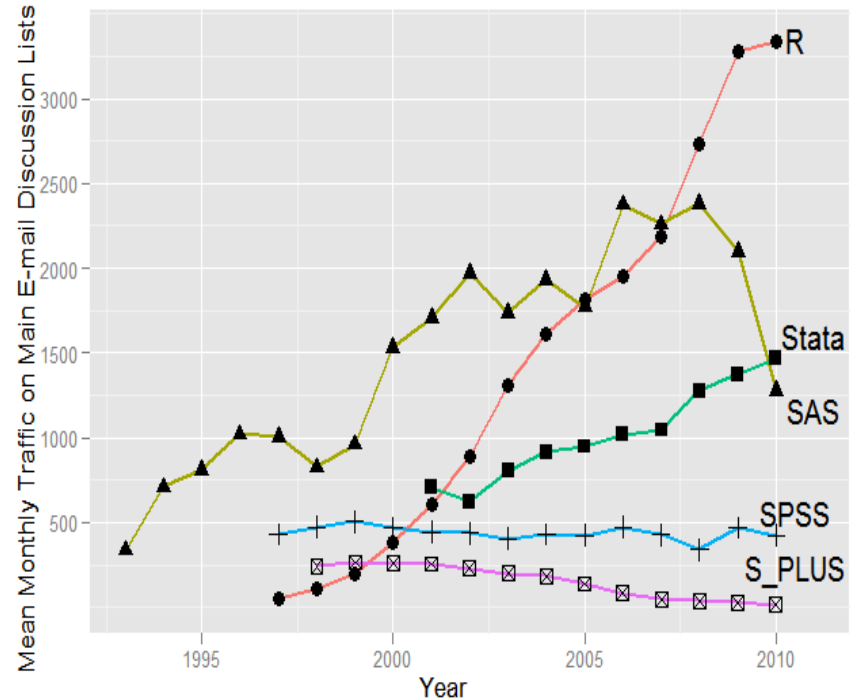
Highly extensible with open source community R packages

Growing Popularity

The R Project for Statistical Computing



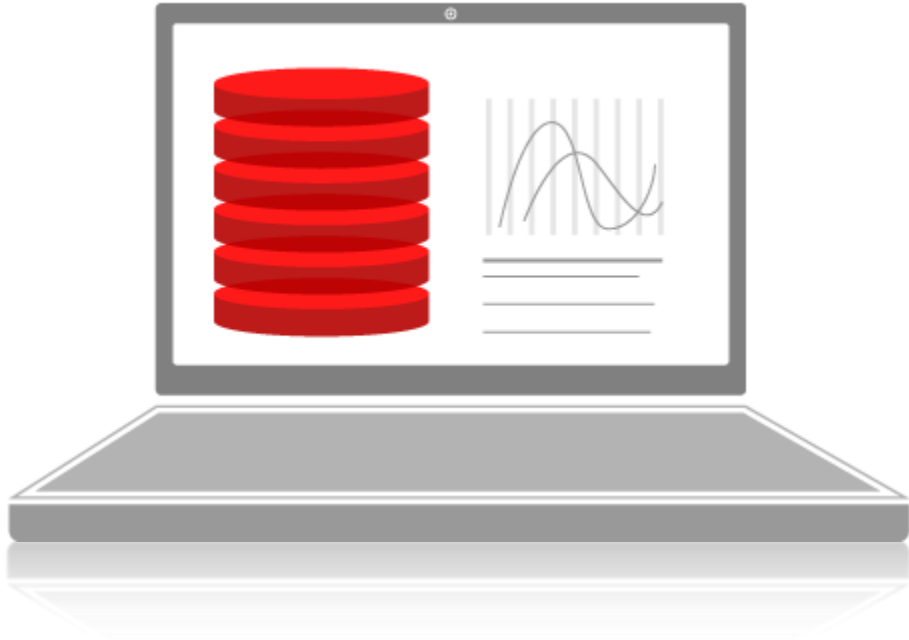
- R's rapid adoption over several years has earned its reputation as a new statistical software standard
 - Rival to SAS and SPSS



While it is difficult to calculate exactly how many people use R, those most familiar with the software estimate that close to 250,000 people work with it regularly.

“[Data Analysts Captivated by R’s Power](#)”, New York Times, Jan 6, 2009

Typical R Approach



Statistical and advanced analyses are run and stored on the user's laptop

What Are R's Challenges?



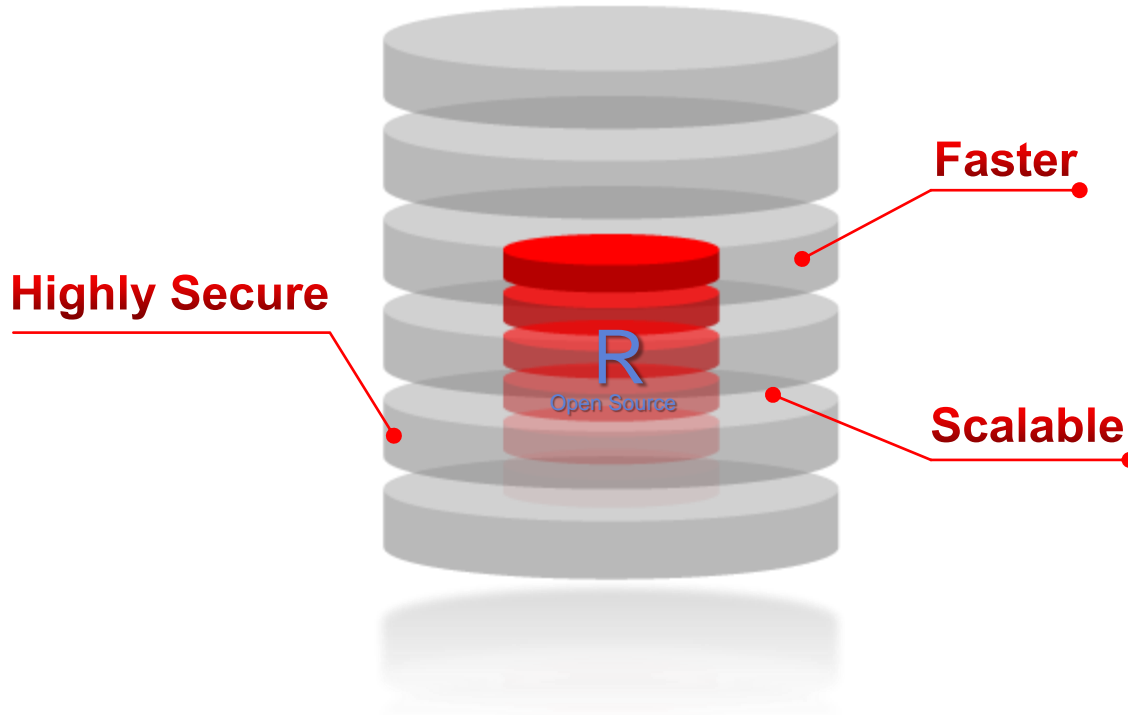
1. R is memory constrained

- R processing is single threaded - does not exploit available compute infrastructure
- R lacks industrial strength for enterprise use cases

2. R has lacked mindshare in Enterprise market

- R is still met with caution by the long established SAS and IBM/SPSS statistical community
 - However, major university (e.g. Yale) Statistics courses now taught in R
 - The FDA has recently shown indications for approval of new drugs for which the submission's data analysis was performed using R

Oracle R Enterprise Approach



Data and statistical analysis are stored and run in-database

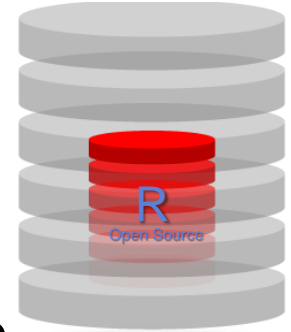
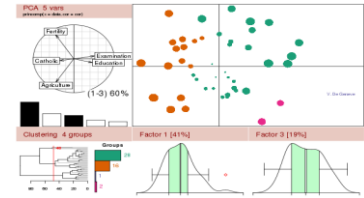
Same R user experience & same R clients

Embed in operational systems

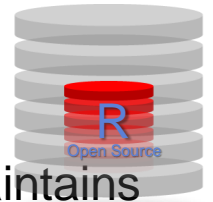
Complements Oracle Data Mining

What is ORACLE® R Enterprise?

- Oracle R Enterprise brings R's statistical functionality closer to the Oracle Database
 1. Eliminate R's memory constraint by enabling R to work directly/transparently on database objects
 - Allows R to run on very large data sets, tables, views
 2. Architected for Enterprise production infrastructure
 - Automatically exploits database parallelism without requiring parallel R programming
 - Build and immediately deploy R scripts



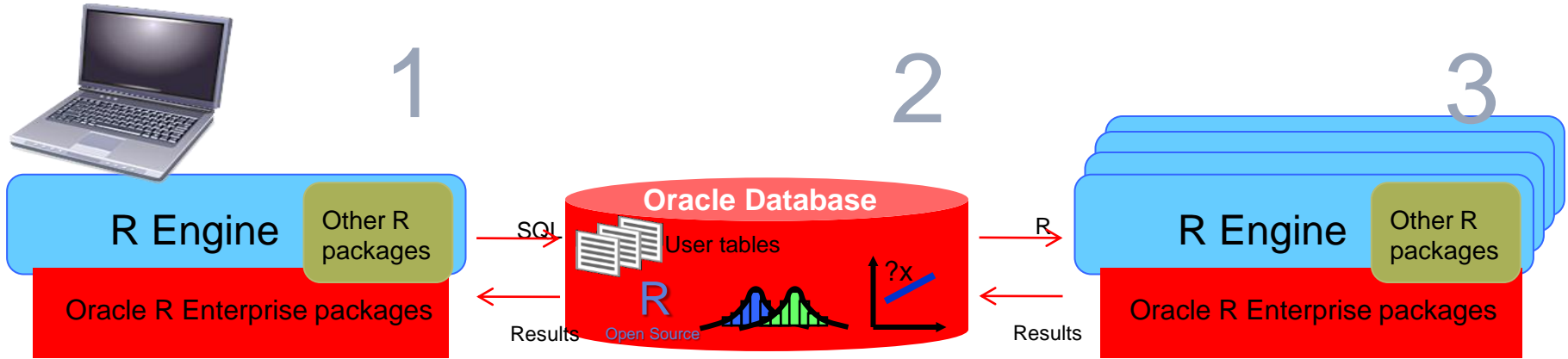
How Oracle R Enterprise Works



ORE Computation Engines

- Oracle R Enterprise eliminates data movement and duplication, maintains security and minimizes latency time from raw data to new information.
 - The database is always involved in serving up data to the R code.
 - Oracle R Enterprise runs in the Oracle Database
- Three ORE Computation Engines
 - Oracle R Enterprise provides three different interfaces between the open-source R engine and the Oracle database:
 1. Oracle R Enterprise (ORE) Transparency Layer
 2. Oracle Statistics Engine
 3. Embedded R

Oracle R Enterprise Compute Engines



User R Engine on desktop

- R-SQL Transparency Framework intercepts R functions for scalable in-database execution
- Submit entire R scripts for execution by Oracle Database

Database Compute Engine

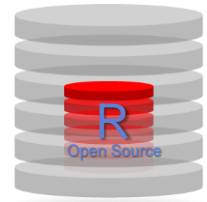
- Access tables, views, and external tables, as well as data through DB LINKS
- Leverage database SQL parallelism
- Leverage new and existing in-database statistical and data mining capabilities

R Engine(s) spawned by Oracle DB

- Database can spawn multiple R engines for database-managed parallelism
- Efficient data transfer to spawned R engines
- Emulate map-reduce style algorithms and applications
- Enables "lights-out" execution of R scripts

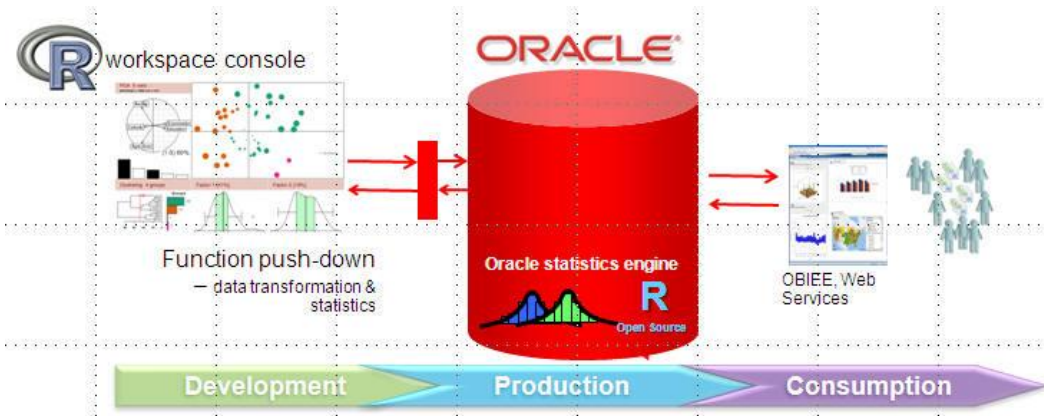
How Oracle R Enterprise Works

ORE Computation Engines



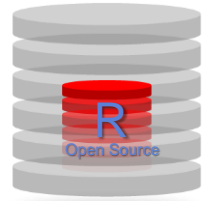
1. Oracle R Enterprise (ORE) Transparency Layer

- Traps all R commands and scripts prior to execution and looks for opportunities to function ship them to the database for native execution
- ORE transparency layer converts R commands/scripts into SQL equivalents and thereby leverages the database as a compute engine.



How Oracle R Enterprise Works

ORE Computation Engines



2. In-Database Statistics Engine

- Significantly extends the Oracle Database's library of statistical functions and advanced analytical computations
- Provides support for the complete R language and statistical functions found in **Base R** and **selected R packages** based on customer usage
 - Open source packages - written entirely in R language with only the functions for which we have implemented SQL counterparts - can be translated to execute in database.
- Without anything visibly different to the R users, their R commands and scripts are oftentimes accelerated by a factor of 10-100x

All Base R functions
R Multiple Regression

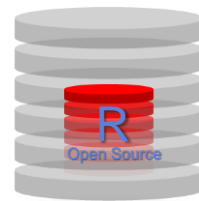
.... Driven by customers

ORE Functions

- ORE SUMMARY
- ORE FREQUENCY
- ORE CORR
- ORE UNIVARITE
- ORE CROSSTAB
- ORE RANK
- ORE SORT
- ...

How Oracle R Enterprise Works

ORE Computation Engines



3. Embedded R Engine

- For R functions not able to be mapped to native in-database functions, Oracle R Enterprise **makes “extproc” remote procedure calls** to multiple R engines running on multiple database servers/nodes
- This Oracle R Enterprise embedded layer uses the database as a data provider providing data level parallelism to R code

Working with ORE Connections

```
ore.connect("RUSER", "SID", "HOST", "PASSWORD", 1521)
```

Connect to a specific schema and database
One connection active at a time

```
ore.create( ONTIME_S, table = "NEW_ONTIME_S")  
ore.create( ONTIME_S, view = "NEW_ONTIME_S_VIEW")
```

Create a database table from a data.frame,
ore.frame. Create a view from an ore.frame.

```
ore.drop(table="NEW_ONTIME_S")  
ore.drop(view="NEW_ONTIME_S_VIEW")
```

Drop table or view in database

```
v <- ore.push(c(1,2,3,4,5))
```

Store R object in database as temporary
object, returns handle to object. Data frame,
matrix, and vector to table, list/model/others
to serialized object

```
ore.sync()  
ore.sync("RUSER")  
ore.sync(table=c("ONTIME_S", "NARROW"))  
ore.sync("RUSER", table=c("ONTIME_S", "NARROW"))
```

Synchronize ORE proxy objects in R with
tables/views available in database, on a per
schema basis

```
ore.exists("ONTIME_S", "RUSER")
```

Returns TRUE if named table or view exists in
schema

Working with ORE Connections

```
ore.ls()
```

```
ore.ls("RUSER")
```

```
ore.ls("RUSER",all.names=TRUE)
```

```
ore.ls("RUSER",all.names=TRUE, pattern= "NAR")
```

List the objects available in ORE environment mapped to database schema.

All.names=FALSE excludes names starting with a '.'

```
t <- ore.get("ONTIME_S","RUSER")
```

Obtain object to named table/view in schema.

```
ore.attach("RUSER")
```

```
ore.attach("RUSER", pos=2)
```

Make database objects visible in R for named schema. Can place corresponding environment in specific position in env path.

```
ore.detach("RUSER")
```

Remove schema's environment from the object search path.

```
ore.rm("DF1")
```

```
ore.rm(list("TABLE1","TABLE2"), "RUSER")
```

Remove table or view from schema's R environment.

```
ore.disconnect()
```

Disconnect from the database. Clean up all associated R objects and temporary database schema objects

```
ore.exec("create table F2 as select * from ONTIME_S")
```

Execute SQL or PL/SQL without return value

Create database tables from data.frames

```
df <- data.frame(A=1:26, B=letters[1:26])  
dim(df)  
class(df)  
ore.create(df, table="TEST_DF")  
ore.ls(pattern="TEST_DF")  
class(TEST_DF)  
dim(TEST_DF)  
head(TEST_DF)  
ore.drop(table="TEST_DF")
```



R user on desktop

Client R Engine

Other R packages

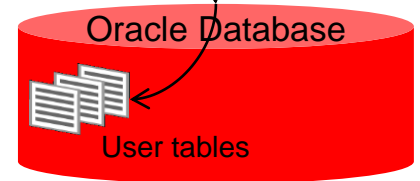
Transparency Layer

Oracle R package

```
R> df <- data.frame(A=1:26, B=letters[1:26])  
R> dim(df)  
[1] 26 2  
R> class(df)  
[1] "data.frame"  
R> ore.create(df, table="TEST_DF")  
R> ore.ls(pattern="TEST_DF")  
[1] "TEST_DF"  
R> class(TEST_DF)  
[1] "ore.frame"  
attr(,"package")  
[1] "OREbase"  
R> dim(TEST_DF)  
[1] 26 2  
R> head(TEST_DF)  
  A B  
0 1 a  
1 2 b  
2 3 c  
3 4 d  
4 5 e  
5 6 f  
R> ore.drop(table="TEST_DF")
```

Goal: Create a database table named "TEST_DF" from an R data.frame object.

Enables user to transition from laptop files to database schema



SQL and Parallel Execution

Embedded Script Execution – SQL Interface

R Interface function	Purpose
<code>rqEval()</code>	Invoke stand-alone R script
<code>rqTableEval()</code>	Invoke R script with full table as input
<code>rqRowEval()</code>	Invoke R script on one row at a time, or multiple rows in chunks
<code>rqGroupEval()</code>	Invoke R script on data partitioned by grouping column
<code>sys.rqScriptCreate</code>	Create named R script
<code>sys.rqScriptDrop</code>	Drop named R script

ORE function	Input data	FUN.VALUE	Arguments	R Script	Special
rqEval()	Internally generated data	NULL (returns chunked blob)	NULL	String argument	Not applicable

```

begin
  sys.rqScriptCreate('Example1',
'function() {
  ID <- 1:10
  res <- data.frame(ID = ID, RES = ID / 100)
  res}');
end;
/
select *
  from table(rqEval(NULL,
    'select 1 id, 1 res from dual',
    'Example1'));

```

- Execute R script with no parameters
- Specify output to return two numbers
 - id
 - res

```

SQL> begin
  sys.rqScriptCreate('Example1',
'function() {
  ID <- 1:10
  res <- data.frame(ID = ID, RES = ID / 100)
  res}');
end;
/
select *
  from table(rqEval(NULL,
    'select 1 id, 1 res from dual',
    'Example1'));
      2  3  4  5  6  7  8
PL/SQL procedure successfully completed.

```

```

SQL>
      2  3  4
      ID  RES
-----
      1   .01
      2   .02
      3   .03
      4   .04
      5   .05
      6   .06
      7   .07
      8   .08
      9   .09
     10   .1

```

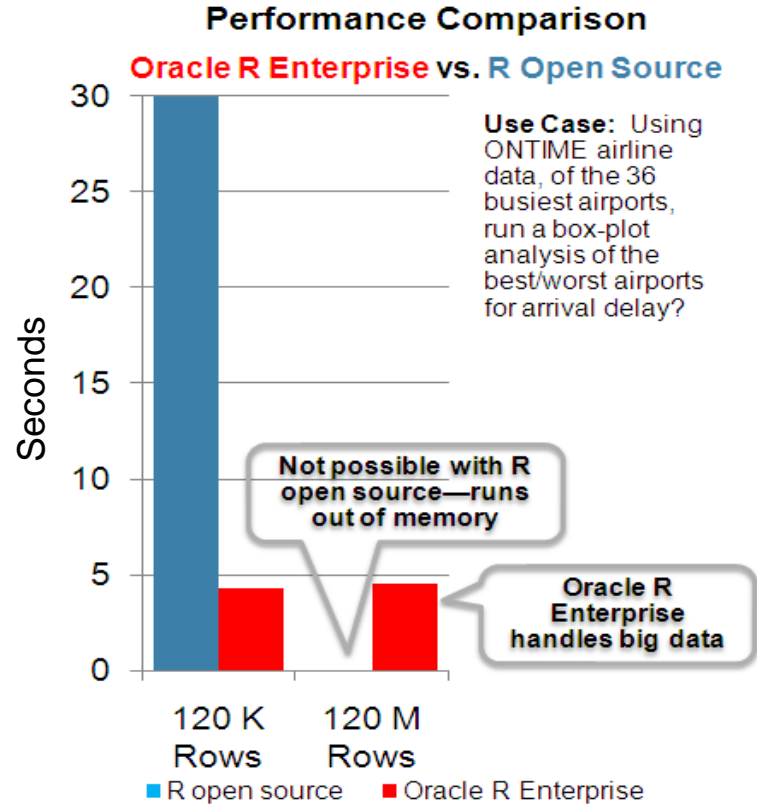
10 rows selected.

Parallelism in the Transparency Layer

- Ideal for “bigger data”,
- Operations performed in-database leverage database parallelism
- Database and table must be configured for parallelism as above

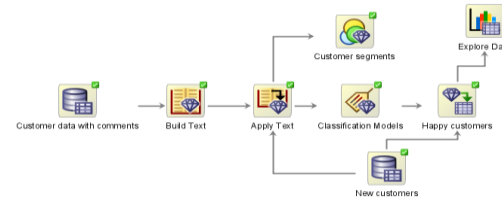
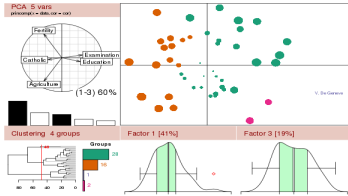
Architecture and Performance

- Transparently function-ships R constructs to database via R → SQL translation
- Performs data-heavy computations in database
 - R for summary analysis and graphics
- Transparent implementation enables using wide range of R “packages” from open source community



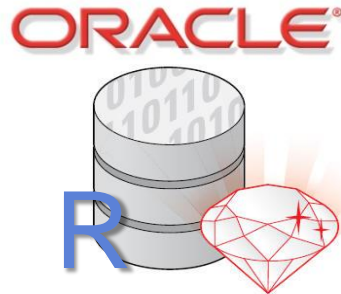
Oracle In-Database Advanced Analytics

Comprehensive Advanced Analytics Platform



Oracle R Enterprise

- Popular open source statistical programming language & environment
- Integrated with database for scalability
- Wide range of statistical and advanced analytical functions
- R embedded in enterprise apps & OBIEE
- Exploratory data analysis
- Extensive graphics
- Open source R (CRAN) packages
- Integrated with Hadoop for HPC



Oracle Data Mining

- Automated knowledge discovery inside the Database
- 12 in-database data mining algorithms
- Text mining
- Predictive analytics applications development environment
- Star schema and transactional data mining
- Exadata "scoring" of ODM models
- SQL Developer/Oracle Data Miner GUI

Statistics

Advanced Analytics

Data & Text Mining

Predictive Analytics

ORACLE

