

# Running R in the Database using Oracle R Enterprise

Brendan Tierney



www.oralytics.com    t : @brendantierney    e : brendan.tierney@oralytics.com



ORACLE ACE Director

- Data Warehousing since 1997
- Data Mining since 1998
- Analytics since 1993



Google brendan tierney data mining



www.oralytics.com    t : @brendantierney    e : brendan.tierney@oralytics.com



## Agenda

- What is R?
- Oracle Advanced Analytics Option
- Oracle R Technologies & Oracle R Enterprise
- Examples of using ORE
- Creating & running R in the Database
- How to run R in the Database using SQL
- Using ORE with other products



## Agenda

- What is R?
- Oracle Advanced Analytics Option
- Oracle R Technologies & Oracle R Enterprise
- Examples of using ORE
- Creating & running R in the Database
- How to run R in the Database using SQL
- Using ORE with other products



www.oralytics.com

t : @brendantierney

e : brendan.tierney@oralytics.com

ORALYTICS

# R

- R Open source statistical computing and graphics language
- Started in 1993 as an alternative to SAS, SPSS and other proprietary statistical packages
  - Originally called S, renamed to R in 1996
- R is a client and server bundled together as one executable
  - It is a single user tool
  - It is not multi-threaded
  - Constrained to a single CPU
- Millions of R users worldwide
- Thousands of libraries available at
  - <http://cran.r-project.org>
- Free



CRAN  
[Menus](#)  
[What's new?](#)  
[Task Views](#)  
[Search](#)

About R  
[R Homepage](#)  
[The R Journal](#)

Software  
[R Sources](#)  
[R Binaries](#)  
[Packages](#)  
[Other](#)

Documentation  
[Manuals](#)  
[FAQs](#)  
[Contributed](#)

#### Milestones:

2018-01-17: 12100 packages  
 2017-06-10: 10793 packages  
 2017-01-09: 9870 packages  
 2016-06-01: 8492 packages  
 2015-03-13: 6400 packages  
 2015-02-15: 6325 packages  
 2014-10-29: 6000 packages  
 2013-11-08: 5000 packages  
 2012-08-23: 4000 packages  
 2011-05-12: 3000 packages  
 2009-10-04: 2000 packages  
 2007-04-12: 1000 packages  
 2004-10-01: 500 packages  
 2003-04-01: 250 packages

#### Available CRAN Packages By Name

ABCDEFGHIJKLMNOPQRSTUVWXYZ

**A3** Accurate, Adaptable, and Accessible Error Metrics for Predictive Models  
**abs** Tools for Approximate Bayesian Computation (ABC)  
**absdEFA** ABCDE\_FBA: A Biologist-Can-Do-Everything of Flux Balance Analysis with this packa  
**ABCExtremes** ABC Extremes  
**ABCoptim** Implementation of Artificial Bee Colony (ABC) Optimization  
**ABCp2** Approximate Bayesian Computational model for estimating P2  
**abtools** Tools for ABC analyses  
**abd** The Analysis of Biological Data  
**abf2** Load Assn ABF2 files (currently only in gap-free recording mode)  
**abind** Combine multi-dimensional arrays  
**abn** Data Modelling with Additive Bayesian Networks  
**abundant** Abundant regression and high-dimensional principal fitted components  
**accionometry** Functions for processing minute-to-minute accelerometer data  
**AcceptanceSampling** Creation and evaluation of Acceptance Sampling Plans  
**ACCLMA** ACC & LMA Graph Plotting  
**accrual** Bayesian Accrual Prediction  
**accrual** Visualization tools for partially accruing data  
**ACD** Categorical data analysis with complete or missing responses  
**acc** Assay-based Cross-sectional Estimation of incidence rates  
**ace()** ace() and avas() for selecting regression transformations  
**acsr** The ACER Method for Extreme Value Estimation  
**aCGH.Spline** Robust spline interpolation for dual color array comparative hybridisation data  
**acsrfit** Align-and-Count Method comparisons of RFLP data  
**ACSE** Affymetrix SNP probe-summatization using non-negative matrix factorization  
**acspia** Modelling dependence with multivariate Archimix (or any user-defined continuous) copula  
**acRM** Convenience functions for analytical Customer Relationship Management  
**acs** Download, manipulate, and present data from the US Census American Community Survey  
**acs** Algorithmic Complexity for Short Strings  
**acs.data** Data Only: Algorithmic Complexity of Short Strings (Computed via Coding Theorem Meth  
**ACTCD** Asymptotic Classification Theory for Cognitive Diagnosis  
**Actigraphy** Actigraphy Data Analysis  
**actuar** Actuarial functions  
**ActuData** Functions for actuarial sciences  
**ada** ada: an R package for stochastic boosting  
**adaboost** Apples multiclass AdaBoost.M1, AdaBoost.SAMME and Bagging

www.oralytics.com

t : @brendantierney

e : brendan.t



- Is used every where
- Particularly in USA
  - And everywhere else



I'm right, and everyone else is wrong.

e : brendan.tierney

```
> library(RJDBC)
> # Create connection driver and open
> connectionjdbcDriver <- JDBC(driverClass="oracle.jdbc.OracleDriver", classPath="c:/ojdbc6.jar")
> jdbcConnection <- dbConnect(jdbcDriver, "jdbc:oracle:thin:@//localhost:1521/orcl", "dmuser", "dmuser")
> #list the tables in the schema
> #dbListTables(jdbcConnection)
> #get the DB connections details - it get LOTS of info - Do not run unless it is really needed
> dbGetInfo(jdbcConnection)
> # Query on the Oracle instance name.
> #instanceName <- dbGetQuery(jdbcConnection, "SELECT instance_name FROM v$instance")
      TABLE_NAME1
1  INSUR_CUST_LTV_SAMPLE2
2      OUTPUT_1_2
> #print(instanceName)tableNames <- dbGetQuery(jdbcConnection, "SELECT table_name from user_tables where
      table_name not like 'DM$%' and table_name not like 'ODMR$%'")
> print(tableNames)
> viewNames <- dbGetQuery(jdbcConnection, "SELECT view_name from user_views")print(viewNames)
1  MINING_DATA_APPLY_V
2  MINING_DATA_BUILD_V
3  MINING_DATA_TEST_V
4  MINING_DATA_TEXT_APPLY_V
5  MINING_DATA_TEXT_BUILD_V
6  MINING_DATA_TEXT_TEST_V
> dbDisconnect(jdbcConnection)
```

Using RJDBC

www.oralytics.com

t : @brendantierney

e : brendan.tierney@oralytics.com

ORALYTICS

```

> library(ROracle)
> drv <- dbDriver("Oracle")
> # Create the connection string
> host <- "localhost"
> port <- 1521
> sid <- "orcl"
> connect.string <- paste(" (DESCRIPTION=", " (ADDRESS=(PROTOCOL=tcp) (HOST=", host, ") (PORT=", port, ") ",
>   "(CONNECT_DATA=(SID=", sid, "))", sep = "")
> con <- dbConnect(drv, username = "dmuser", password = "dmuser", dbname=connect.string)
> rs <- dbSendQuery(con, "select view_name from user_views")
> # fetch records from the resultSet into a data.frame
> data <- fetch(rs)
> # extract all rows
> dim(data)
[1] 6 1
> data
      VIEW_NAME
1  MINING_DATA_APPLY_V
2  MINING_DATA_BUILD_V
3  MINING_DATA_TEST_V
4  MINING_DATA_TEXT_APPLY_V
5  MINING_DATA_TEXT_BUILD_V
6  MINING_DATA_TEXT_TEST_V
> dbCommit(con)
> dbClearResult(rs)
> dbDisconnect(con)

```

Using ROracle

Needs Oracle Client  
in the search pathPulls the data to the  
ClientHas a set of R  
functions tuned for  
the Oracle DB

www.oralytics.com

t : @brendantierney

e : brendan.tierney@oralytics.com

ORALYTICS



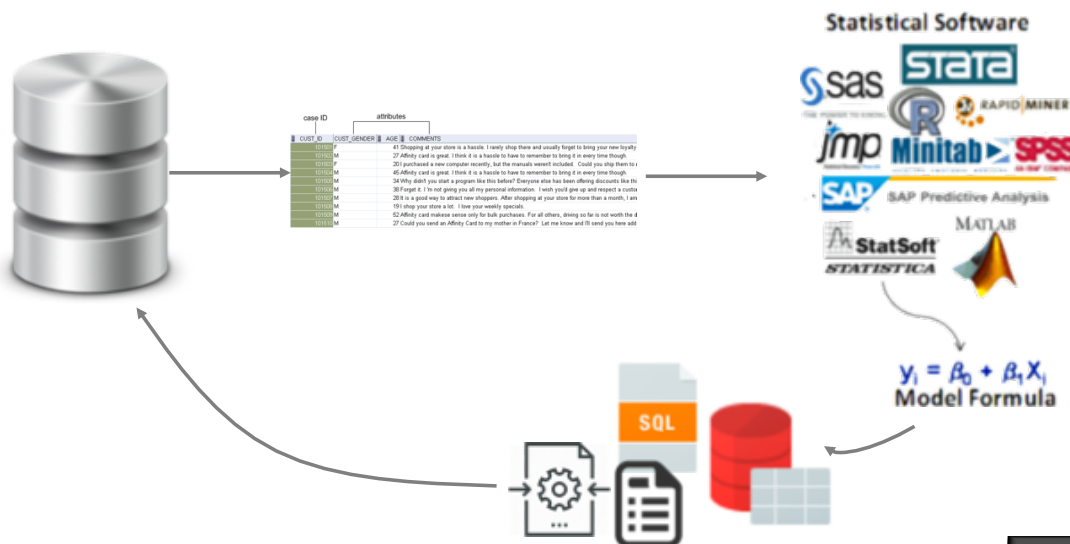
www.oralytics.com

t : @brendantierney

e : brendan.tierney@oralytics.com

ORALYTICS

## Traditional Approach



www.oralytics.com

t : @brendantierney

e : brendan.tierney@oralytics.com

ORALYTICS

## The Challenges

- Scalability
  - Regardless of the number of cores on your CPU, R will only use 1 on a default build
- Performance
  - R reads data into memory by default. Easy to exhaust RAM by storing unnecessary data. Typically R will throw an exception at 2GB.
  - Parallelization can be challenge. Is not Default. Packages available
- Production Deployment
  - Difficulties deploying R in production
  - Typically need to re-code in .....

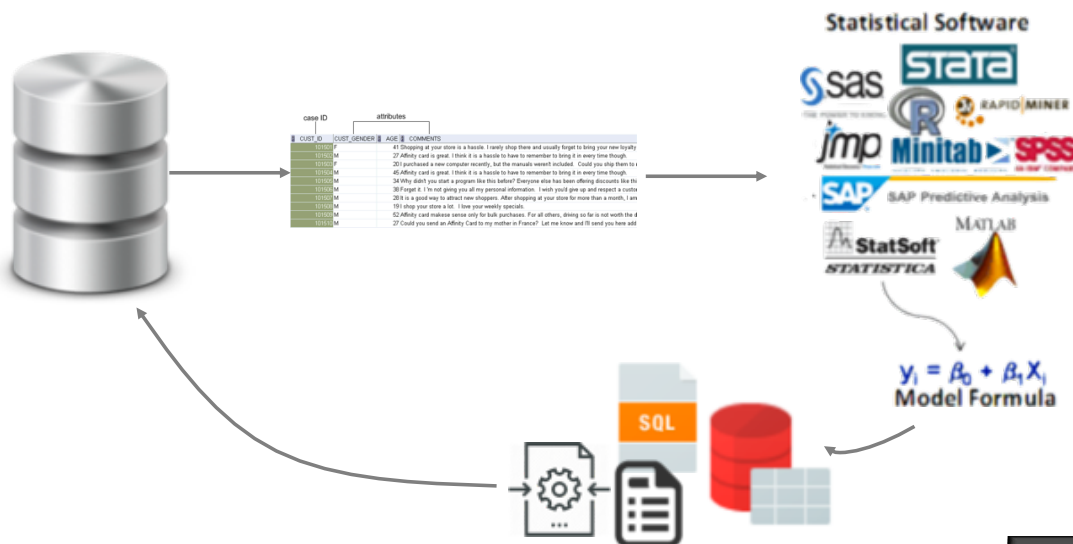
www.oralytics.com

t : @brendantierney

e : brendan.tierney@oralytics.com

ORALYTICS

## Database Centric Approach

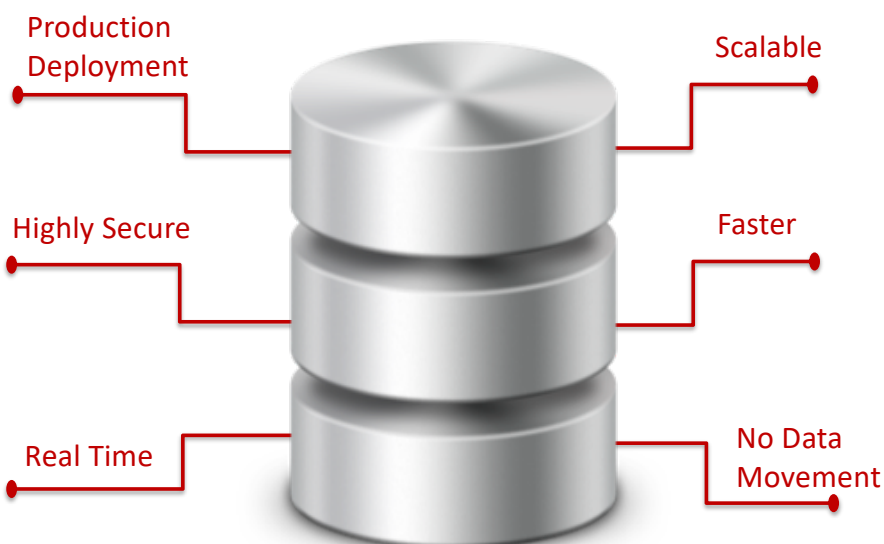


www.oralytics.com

t : @brendantierney

e : brendan.tierney@oralytics.com

ORALYTICS



www.oralytics.com

t : @brendantierney

e : brendan.tierney@oralytics.com

ORALYTICS



## Agenda

- What is R?
- Oracle Advanced Analytics Option
- Oracle R Technologies & Oracle R Enterprise
- Examples of using ORE
- Creating & running R in the Database
- How to run R in the Database using SQL
- Using ORE with other products

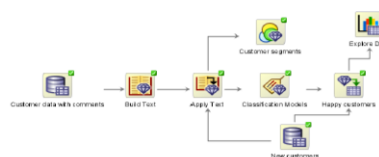
www.oralytics.com

t : @brendantierney

e : brendan.tierney@oralytics.com

ORALYTICS

## Comprehensive Advanced Analytics Platform



### Oracle Data Mining

- **SQL** kernel; automated knowledge discovery inside the Database
- **18** 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

www.oralytics.com

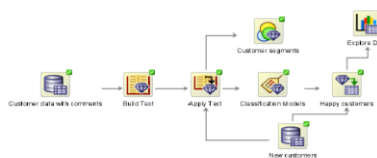
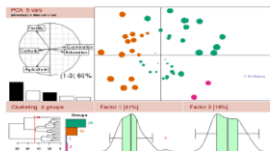
t : @brendantierney

e : brendan.tierney@oralytics.com

ORALYTICS



# Comprehensive Advanced Analytics Platform



## Oracle R Enterprise

- Popular open source R 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

- SQL kernel; automated knowledge discovery inside the Database
- 18 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

# Oracle R Technologies



<p>R Distribution</p> 	<p>Oracle's supported redistribution of open source R, provided as a free download from Oracle, enhanced with dynamic loading of high performance linear algebra libraries.</p>
<p>Oracle R Enterprise</p> 	<p>Integration of R with Oracle Database. A component of the Oracle Advanced Analytics Option. Oracle R Enterprise makes the open source R statistical programming language and environment ready for the enterprise with scalability, performance, and ease of production deployment.</p>
<p>Oracle R Advanced Analytics for Hadoop</p> 	<p>High performance native access to the Hadoop Distributed File System (HDFS) and MapReduce programming framework for R users. Oracle R Advanced Analytics for Hadoop is a component of Oracle Big Data Connectors software suite.</p>
<p>ROracle</p> 	<p>An open source R package, maintained by Oracle and enhanced to use the Oracle Call Interface (OCI) libraries to handle database connections - providing a high-performance, native C-language interface to Oracle Database.</p>

# What has Oracle done?



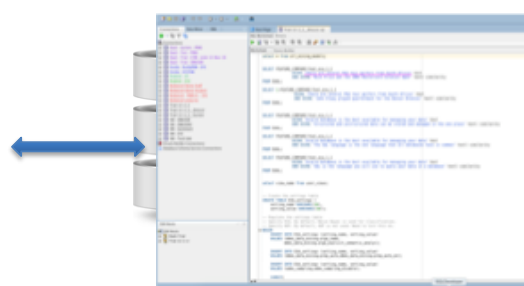
www.oralytics.com

t : @brendantierney

e : brendan.tierney@oralytics.com

ORALYTICS

# What has Oracle done?



www.oralytics.com

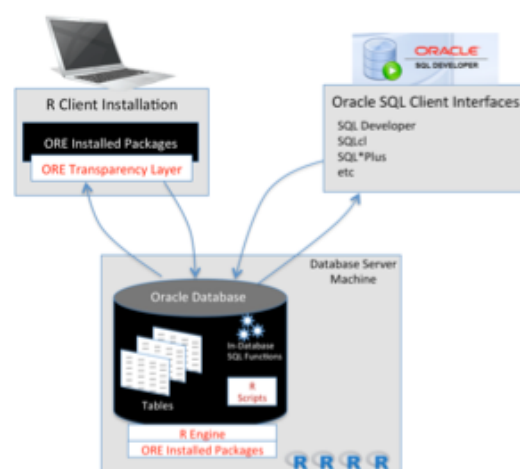
t : @brendantierney

e : brendan.tierney@oralytics.com

ORALYTICS

## Oracle R Enterprise

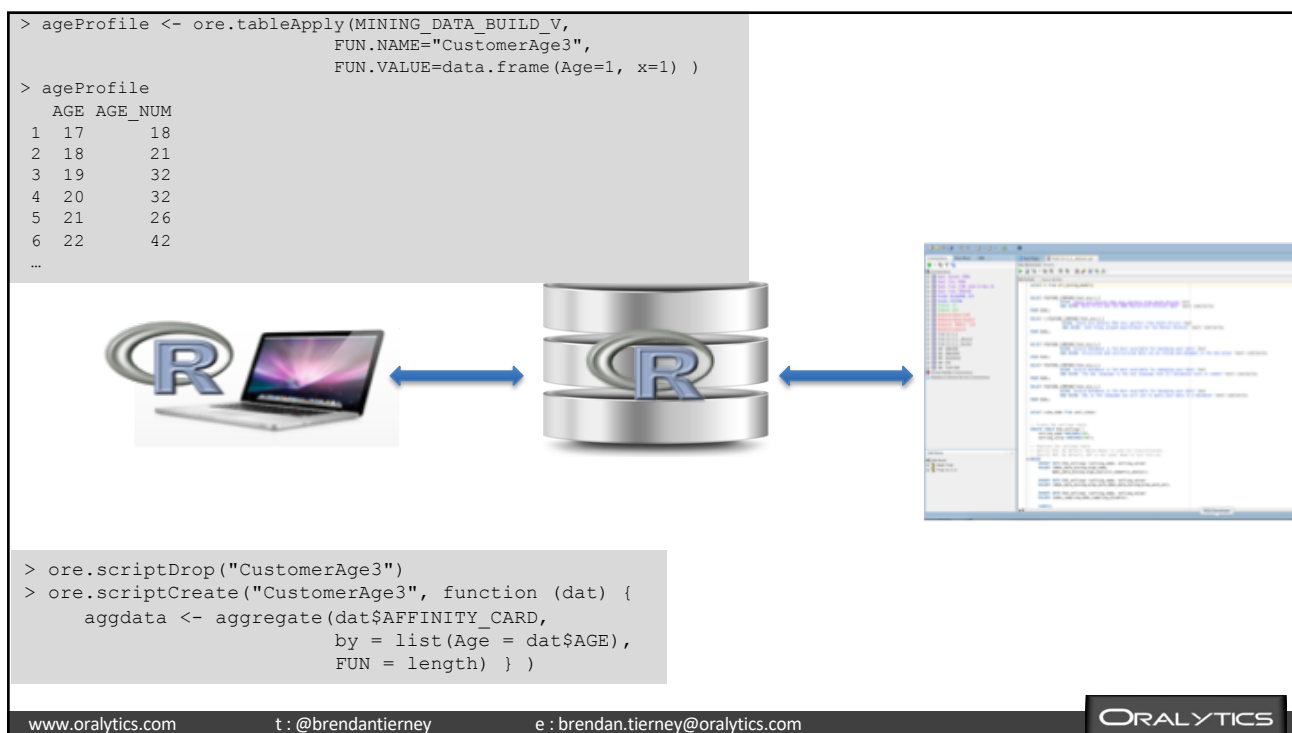
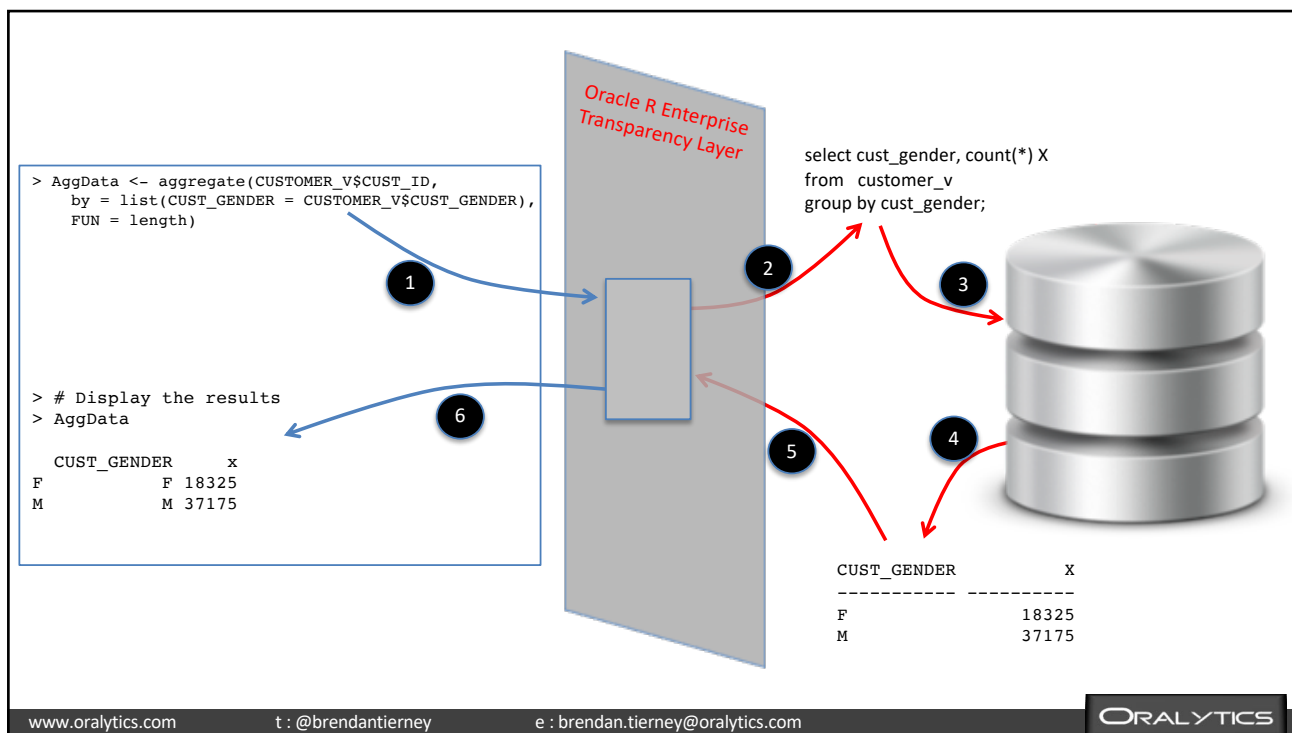
- R installed in ORACLE\_HOME
  - Fully integrated with the database
  - Overcomes the limitations of R
    - Utilizes the DB performance and scalability features
- Full integration into the DB engine
    - Can run R inside the DB
    - Can store R object in the DB
    - Can run R objects using SQL & PL/SQL
    - Easily integrated into other Oracle Tools and Applications
- Greatly expands the statistics & analytics
    - Easily integrates new “bells & whistles” package comes available

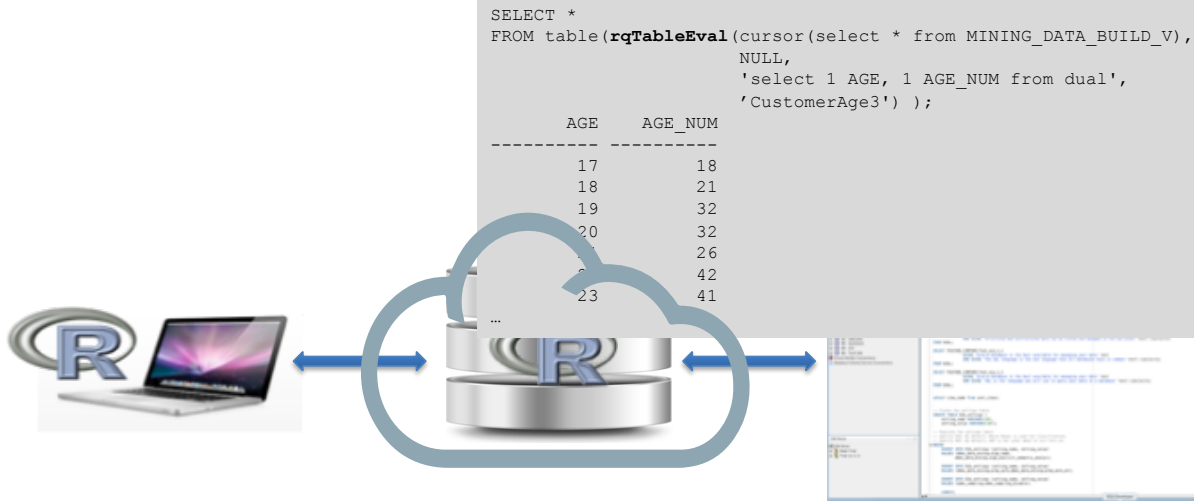


## The Magic : The Transparency Layer

- No need to learn a different programming paradigm or environment
  - If you are an R programmer
- Operate on database data as though they were R objects using R syntax
- Require minimal change to base R scripts for database data
- Implicitly translates R to SQL for in-database execution, performance, and scalability

*The Transparency Layer supports in-database data exploration, data preparation, and data analysis en route performing predictive analytics with a mix of in-database and CRAN techniques.*





```

SELECT *
FROM table (rqTableEval (cursor (select * from MINING_DATA_BUILD_V,
                                NULL,
                                'select 1 AGE, 1 AGE_NUM from dual',
                                'CustomerAge3') ));

```

AGE	AGE_NUM
17	18
18	21
19	32
20	32
23	26
23	42
23	41

```

> ore.scriptDrop("CustomerAge3")
> ore.scriptCreate("CustomerAge3", function (dat) {
  aggdata <- aggregate(dat$AFFINITY_CARD,
    by = list(Age = dat$AGE),
    FUN = length) })

```

SQL

www.oralytics.com    t : @brendantierney    e : brendan.tierney@oralytics.com    ORALYTICS

## The Magic : The Transparency Layer

### ▪ R interface to in-Database Statistical functions

- Special Functions
  - Gamma function
  - Natural logarithm of the Gamma function
  - Digamma function
  - Trigamma function
  - Error function
  - Complementary error function
- Tests
  - Chi-square, McNemar, Bowker
  - Simple and weighted kappas
  - Cochran-Mantel-Haenzel correlation
  - Cramer's V
  - Binomial, KS, t, F, Wilcox
- Base SAS equivalents
  - Freq, Summary, Sort
  - Rank, Corr, Univariate
- Density, Probability, and Quantile Functions
  - Beta distribution
  - Binomial distribution
  - Cauchy distribution
  - Chi-square distribution
  - Exponential distribution
  - F-distribution
  - Gamma distribution
  - Geometric distribution
  - Log Normal distribution
  - Logistic distribution
  - Negative Binomial distribution
  - Normal distribution
  - Poisson distribution
  - Sign Rank distribution
  - Student's t distribution
  - Uniform distribution
  - Weibull distribution
  - Density Function
  - Probability Function
  - Quantile

## The Magic : The Transparency Layer

- R interface to in-Database Predictive Analytics functions

### OREdm

- Support Vector Machine
- GLM
- k-Means clustering
- OCluster clustering
- Naïve Bayes
- Decision Trees
- Association Rules
- Attribute Importance

### OREmodels

- Neural Networks
- Linear Regression
- Stepwise Regression
- Generalized Linear Model

www.oralytics.com

t : @brendantierney

e : brendan.tierney@oralytics.com

ORALYTICS

## Agenda

- What is R?
- Oracle Advanced Analytics Option
- Oracle R Technologies & Oracle R Enterprise
- Examples of using ORE
- Creating R objects in the Database
- How to run R in the Database using SQL
- Using ORE with other products



You can run all the example code if you have the ODM Demo Schema created

www.oralytics.com

t : @brendantierney

e : brendan.tierney@oralytics.com

ORALYTICS



www.oralytics.com

t : @brendantierney

e : brendan.tierney@oralytics.com

ORALYTICS

## Embedded R Execution using SQL

SQL API	Equivalent ORE function	Description
rqEval	<b>ore.doEval</b>	Executes a user defined R script that is passed to it and returns any result generated.
rqTableEval	<b>ore.tableApply</b>	Executes a function or script on all rows on a supplied data set.
<i>"rqGroupEval"</i>	<b>ore.groupApply</b>	<p>Executes a function or script on partitions of the supplied data set. The partitions are defined on one or more attributes of the data set. Parallel execution is supported for each partition.</p> <p>NOTE: There is no specific function called <i>"rqGroupEval"</i>. Instead you have to define this function, in a specific way, for your data.</p>
rqRowEval	<b>ore.rowApply</b>	Executes a function or script on a defined set of rows (chunks) from the supplied data set. Parallel execution is supported for each set of rows (chunks).

www.oralytics.com

t : @brendantierney

e : brendan.tierney@oralytics.com

ORALYTICS



```
--
-- Now let us use the Demo_GLM_Batch script to score data in Real-Time
-- The data values are passed to the GLM model
--
select * from table(rqTableEval(
  cursor(select 'M' CUST_GENDER,
             23 AGE,
             'Married' CUST_MARITAL_STATUS,
             'United States of America' COUNTRY_NAME,
             'B: 30,000 - 49,999' CUST_INCOME_LEVEL,
             'Assoc-A' EDUCATION,
             '3' HOUSEHOLD_SIZE,
             5 YRS_RESIDENCE
          from dual),
  cursor(select 'myDatastore' "datastore_name", 1 "ore.connect" from dual),
  'select CUST_GENDER, AGE, CUST_MARITAL_STATUS, COUNTRY_NAME, CUST_INCOME_LEVEL, EDUCATION,
HOUSEHOLD_SIZE, YRS_RESIDENCE, 1 PRED from MINING_DATA_APPLY',
  'Demo_GLM_Batch')) order by 1, 2, 3;
```

What-if analysis

You can easily use this  
in-Database R code in  
your applications

```
8 order by 1, 2, 3:
C      AGE CUST_MARITAL_STATUS  COUNTRY_NAME
-----
CUST_INCOME_LEVEL      EDUCATION      HOUSEHOLD_SIZE      YRS_RESIDENCE      PRED
-----
M      23 Married              United States of America
B: 30,000 - 49,999      Assoc-A              3              5 1.00174544
DMUSER >
```

www.oralytics.com

t : @brendantierney

e : brendan.tierney@oralytics.com

ORALYTICS

## Agenda

- What is R?
- Oracle Advanced Analytics Option
- Oracle R Technologies & Oracle R Enterprise
- Examples of using ORE
- Creating & running R in the Database
- How to run R in the Database using SQL
- Using ORE with other products

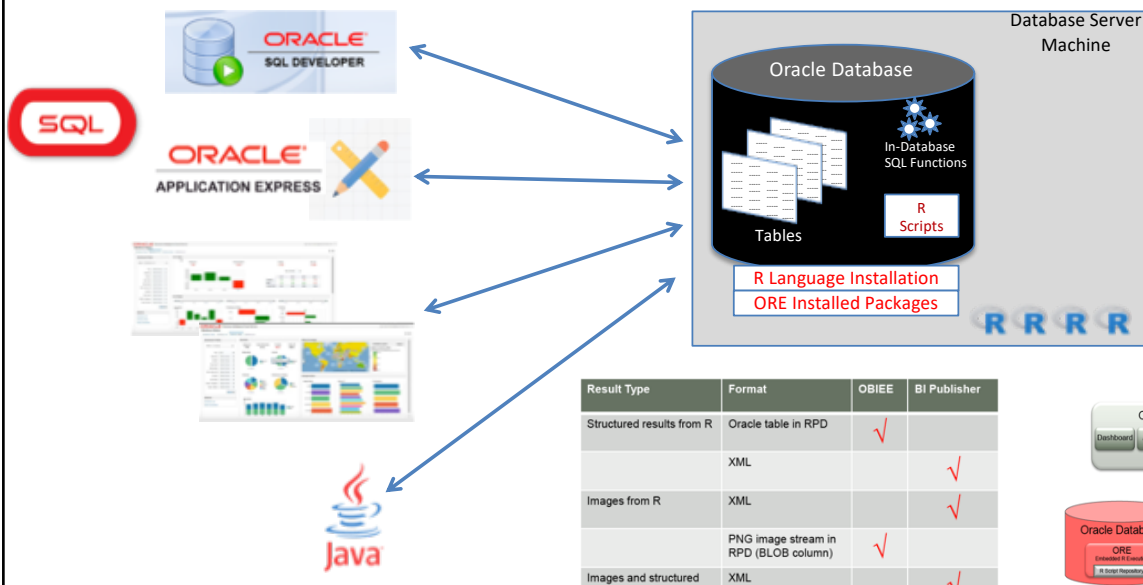
www.oralytics.com

t : @brendantierney

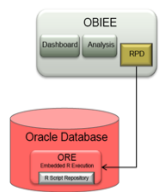
e : brendan.tierney@oralytics.com

ORALYTICS

## Integrating with OBIEE, BI Publisher & Any other Language or Tool



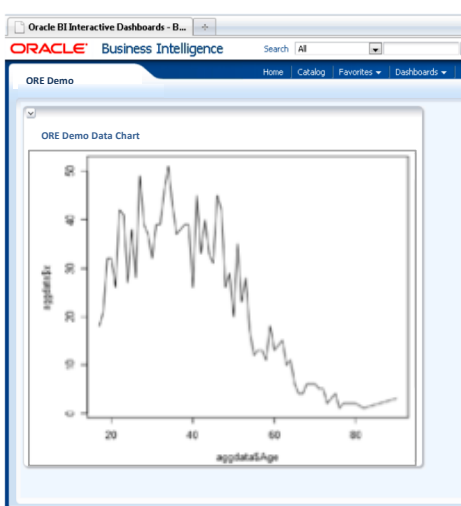
Result Type	Format	OBIEE	BI Publisher
Structured results from R	Oracle table in RPD	✓	
	XML		✓
Images from R	XML		✓
	PNG image stream in RPD (BLOB column)	✓	
Images and structured results from R	XML		✓



```

-- Create an embedded R script
-- Called using the ORE SQL API
-- - performs an aggregation of the data
-- - creates a Graphic Plot
--
begin
  sys.rqScriptDrop('AgeProfile');
  sys.rqScriptCreate('AgeProfile',
    'function(dat) {
      mdbv <- dat
      aggdata <- aggregate(mdbv$AFFINITY_CARD,
                          by = list(Age = mdbv$AGE),
                          FUN = length)
      res <- plot(aggdata$Age, aggdata$x, type = "l") } ');
end;
/
--
-- Execute the embedded R Script
-- - Graphic created in PNG format for import into OBIEE
-- - change PNG to XML for BI Publisher
--
select * from table(rqTableEval( cursor(select * from MINING_DATA_BUILD_V),
                                cursor(select 1 "ore.connect" from dual),
                                'PNG','AgeProfile'));

```



NAME	ID	IMAGE
1	89504E470D001B00000000049484452000001E0000001E0080600000070D4BE950000200049444154789CE0D0797494F58D	9FF0F793054216322108181609C90448D158542C138522B46CB5866CD054

ORE in APEX

Demo Oracle R Enterprise

Report WordCloud

Age Age Num

17	18
18	21
19	32
20	32
21	26
22	42
23	41
24	27
25	36
26	28
27	49
28	39
29	37
30	32
31	39

MAKES

- Accord
- BMW
- Chevrolet
- Dodge
- Ford
- Honda
- Nissan
- Subaru
- Toyota
- VW

<http://www.oralytics.com/2016/07/creating-ggplot2-graphics-using-sql.html>

www.oralytics.com t : @brendantierney e : brendan.tierney@oralytics.com ORALYTICS

## The Challenges : With ORE “Yes we can”

### Scalability

- Regardless of the number of cores on your CPU, R will only use 1 on a default build

Scale with the Database

### Performance

- R reads data into memory by default. Easy to exhaust RAM by storing unnecessary data. Typically R will throw an exception at 2GB.
- Parallelization can be challenge. Is not Default. Packages available

### Production Deployment

- Difficulties deploying R in production
- May need to re-code in .....

SQL

In-Database R execution.

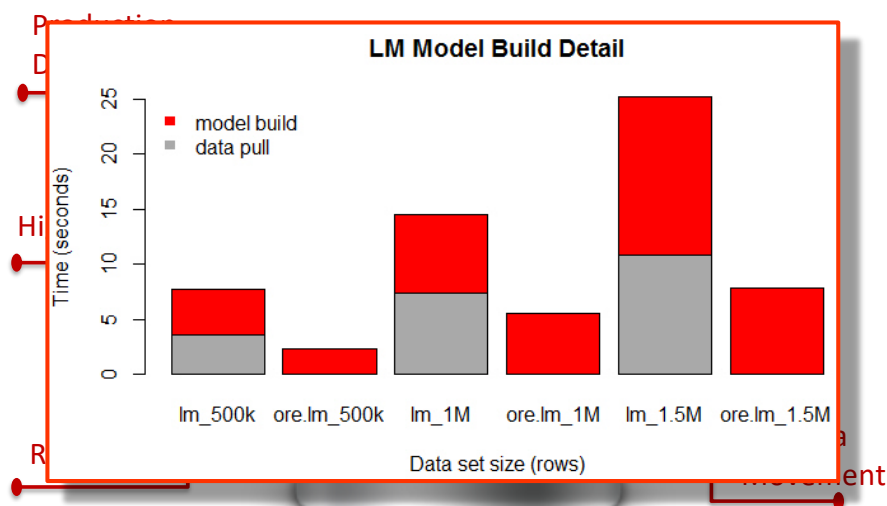
Easy integration with all your applications.

## The Challenges : With ORE “Yes we can”

What if I want to use a new R Package?

Easy, just install it on the DB server and off you go !!!

your applications.



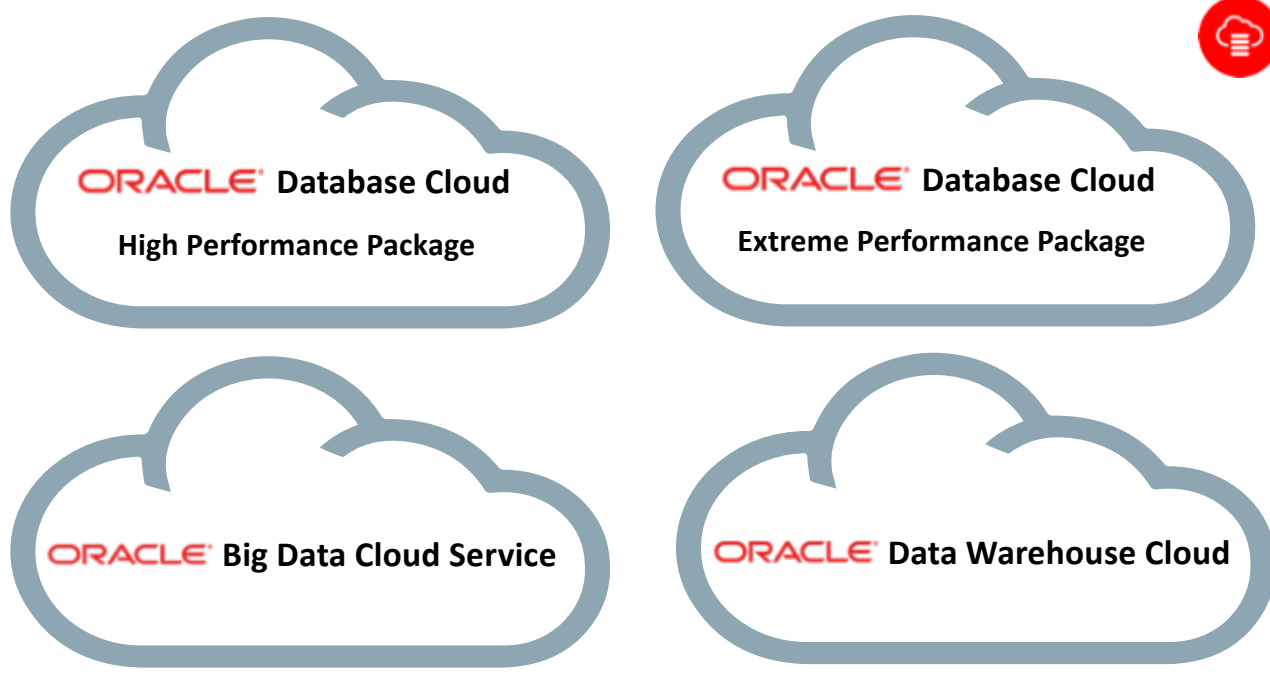
# Data Mining / Data Science in Oracle



Is Just **SQL** (Oracle Data Mining  
& SQL Statistics functions)

+ **R** (Oracle R Enterprise)





**ORACLE® Database Cloud**  
High Performance Package


**ORACLE® Database Cloud**  
Extreme Performance Package

**ORACLE® Big Data Cloud Service**

**ORACLE® Data Warehouse Cloud**

www.oralytics.com    t : @brendantierney    e : brendan.tierney@oralytics.com    ORALYTICS

Google    brendan tierney data mining



@ brendan.tierney@oralytics.com

@brendantierney

www.oralytics.com

ie.linkedin.com/in/brendantierney



**Predictive Analytics Using Oracle Data Miner**  
Oracle® and Java™ Data Mining Models in Oracle, Hadoop, and PL/SQL.  
Brendan Tierney

**Oracle R Enterprise**  
Harnessing the Power of R in the Oracle Database  
Oracle and Extreme Powerful, Real-Time Analytics Using Oracle In-Database Capabilities of R  
Brendan Tierney

**Real World SQL & PL/SQL**  
Advice from the Experts  
Alec Beaulieu, Brendan Tierney, Matt Halsey, Alan Harkin, Martin Whelan

**DATA SCIENCE**  
John Kallman, Brendan Tierney

www.oralytics.com    t : @brendantierney    e : brendan.tierney@oralytics.com    ORALYTICS

Word Cloud of the Oracle Advanced Analytics web-pages



<http://www.oralytics.com/2015/01/creating-word-cloud-of-oracle-aaa.html>

[www.oralytics.com](http://www.oralytics.com)

t : @brendantierney

e : [brendan.tierney@oralytics.com](mailto:brendan.tierney@oralytics.com)

ORALYTICS

# 500+ Technical Experts Helping Peers Globally

ORACLE®

ACE PROGRAM



ORACLE®  
ACE Director



ORACLE®  
ACE



ORACLE®  
ACE Associate

### 3 Membership Tiers

- Oracle ACE Director
- Oracle ACE
- Oracle ACE Associate

[bit.ly/OracleACEProgram](http://bit.ly/OracleACEProgram)

### Connect:

✉ [oracle-ace\\_ww@oracle.com](mailto:oracle-ace_ww@oracle.com)

f Facebook.com/oracleaces

t @oracleace



Nominate yourself or someone you know: [acenomination.oracle.com](http://acenomination.oracle.com)



**ORALYTICS**

Search

**BROWSE DESIGNS**

Shop for one-of-a-kind gifts  
Most orders ship in 24 hours!

Thanks for visiting my online shop! Find what you're looking for yourself or great gifts for your friends. You'll find unique merchandise with my art on t-shirts, sweatshirts, mugs, stickers, and more.

 Sqli Tuning T-Shirt \$20.00	 Sqli Optimizer Tuning T-Shirt \$20.00	 Sqli: The One Language To Rule All Your T-Shirt \$26.00
 MORE COLORS AVAILABLE Sqli: The One Language To Rule All Your T-Shirt \$26.00	 Sqli Organic Men's T-Shirt \$26.00	 MORE COLORS AVAILABLE Sqli Organic Men's T-Shirt (dark) \$26.00
 Sqli Is King Polo Golf Shirt \$22.00	 MORE COLORS AVAILABLE P/sqli Developer T-Shirt (dark) \$20.00	 P/sqli Developer White T-Shirt \$20.00

www.oralytics.com    t : @brendantierney    e : brendan.tierney@oralytics.com    **ORALYTICS**

<http://www.cafepress.com/oralytics>

<https://youtu.be/ZdHK2BWnVD4>

YouTube NL    Search

**Combining the Power of R  
& in-Database Data Mining:  
Running R in the Database. Seriously!!!**

Brendan Tierney  
**ORALYTICS**

0:02 / 50:18

Analytics    Video Manager

www.oralytics.com    t : @brendantierney    e : brendan.tierney@oralytics.com    **ORALYTICS**