Session 1: Introduction to Oracle's R Technologies

Mark Hornick, Director, Oracle Advanced Analytics Development
Oracle Advanced Analytics

Topics

- What is R?
- Oracle R Enterprise motivation and overview
- Oracle R Distribution
- ROracle
- Oracle R Advanced Analytics for Hadoop
- Next level view – Oracle R Enterprise
- Oracle Advanced Analytics option
  - Oracle R Enterprise
  - Oracle Data Mining
- Summary

What is R?

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

- **Started in 1994 as an Alternative to SAS, SPSS & Other proprietary Statistical Environments**
- **The R environment**
  - R is an integrated suite of software facilities for data manipulation, calculation and graphical display
- **Around 2 million R users worldwide**
  - Widely taught in Universities
  - Many Corporate Analysts and Data Scientists know and use R
- **Thousands of open sources packages to enhance productivity such as:**
  - Bioinformatics with R
  - Spatial Statistics with R
  - Financial Market Analysis with R
  - Linear and Non Linear Modeling

CRAN
Mirrors
What's new?
Task Views
Search

About R
R Homepage
The R Journal

Software
R Sources
R Binaries
Packages
Other

Documentation
Manuals
FAQs
Contributed

CRAN Task Views

Bayesian	Bayesian Inference
ChemPhys	Chemometrics and Computational Physics
ClinicalTrials	Clinical Trial Design, Monitoring, and Analysis
Cluster	Cluster Analysis & Finite Mixture Models
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
gR	gRaphical Models in R
HighPerformanceComputing	High-Performance and Parallel Computing with R
MachineLearning	Machine Learning & Statistical Learning
MedicalImaging	Medical Image 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
Survival	Survival Analysis
TimeSeries	Time Series Analysis

CRAN Task View – Machine Learning & Statistical Learning

cran.r-project.org/web/views/MachineLearning.html

Most Visited | Getting Started | Latest Headlines | devops | OOW ORE Keynote | Summary Bug Report | AIME -- A Place for In... | Bookmark

CRAN Task View: Machine Learning & Statistical Learning

Maintainer: Torsten Hothorn
Contact: Torsten.Hothorn at R-project.org
Version: 2011-12-20

Several add-on packages implement ideas and methods developed at the borderline between computer science and statistics - this field of research is usually referred to as machine learning. The packages can be roughly structured into the following topics:

- *Neural Networks* : Single-hidden-layer neural network are implemented in package nnet (shipped with base R). Package RSNNS offers an interface to the Stuttgart Neural Network Simulator (SNNS).
- *Recursive Partitioning* : Tree-structured models for regression, classification and survival analysis, following the ideas in the CART book, are implemented in rpart (shipped with base R) and tree. Package rpart is recommended for computing CART-like trees. A rich toolbox of partitioning algorithms is available in Weka , package RWeka provides an interface to this implementation, including the J4.8-variant of C4.5 and M5. The Cubist package fits rule-based models (similar to trees) with linear regression models in the terminal leaves, instance-based corrections and boosting.

  Two recursive partitioning algorithms with unbiased variable selection and statistical stopping criterion are implemented in package party. Function `ctree()` is based on non-parametrical conditional inference procedures for testing independence between response and each input variable whereas `mob()` can be used to partition parametric models. Extensible tools for visualizing binary trees and node distributions of the response are available in package party as well.

  An adaptation of rpart for multivariate responses is available in package mvpart. A tree algorithm fitting nearest neighbors in each node is implemented in package knnTree. For problems with binary input variables the package LogicReg implements logic regression. Graphical tools for the visualization of trees are available in packages maptree and pinktoe. An approach to deal with the instability problem via extra splits is available in package TWIX.

  Trees for modelling longitudinal data by means of random effects are offered by packages REEMtree and longRPart and trees tailored for ordinal responses by package rpartOrdinal. Partitioning of mixed models is performed by RPMM.

  Computational infrastructure for representing trees and unified methods for predition and visualization is implemented in partykit. This

- ahaz
- arules
- BayesTree
- Boruta
- BPHO
- bst
- caret
- CORElearn
- CoxBoost
- Cubist
- e1071 (core)
- earth
- elasticnet
- ElemStatLearn
- evtree
- gafit
- GAMBoost
- gamboostLSS
- gbev
- gbm (core)
- glmnet
- glmpath
- GMMBoost
- grplasso
- hda
- ipred
- kernlab (core)
- klaR
- lars
- lasso2
- LiblineaR
- LogicForest
- LogicReg
- longRPart
- mboost (core)
- mvpart
- ncvreg
- nnet (core)
- oblique.tree
- obliqueRF
- pamr
- party
- partykit
- penalized
- penalizedSVM
- predbayescor
- quantregForest
- randomForest (core)
- randomSurvivalForest
- rattle
- rda
- rdetools
- REEMtree
- relaxo
- rgenoud
- rgp
- rminer
- ROCR
- rpart (core)
- rpartOrdinal
- RPMM
- RSNNS
- RWeka
- sda
- SDDA
- svmpath
- tgp
- tree
- TWIX
- varSelRF

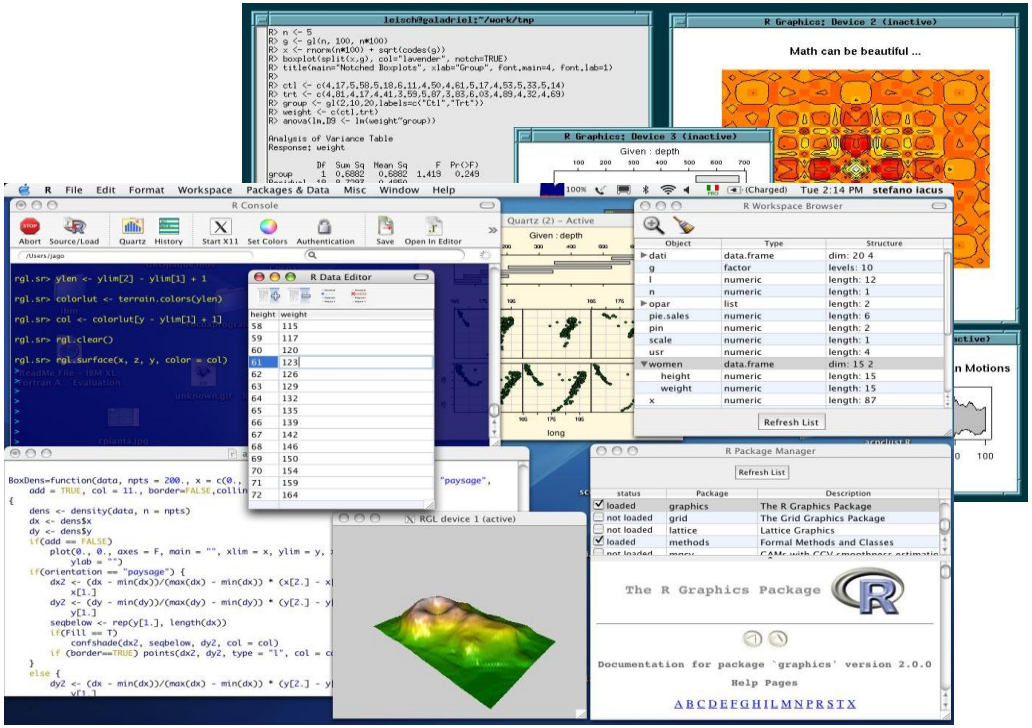
Why statisticians | data analysts | data scientists use R

R is a statistics language similar to Base SAS or SPSS statistics

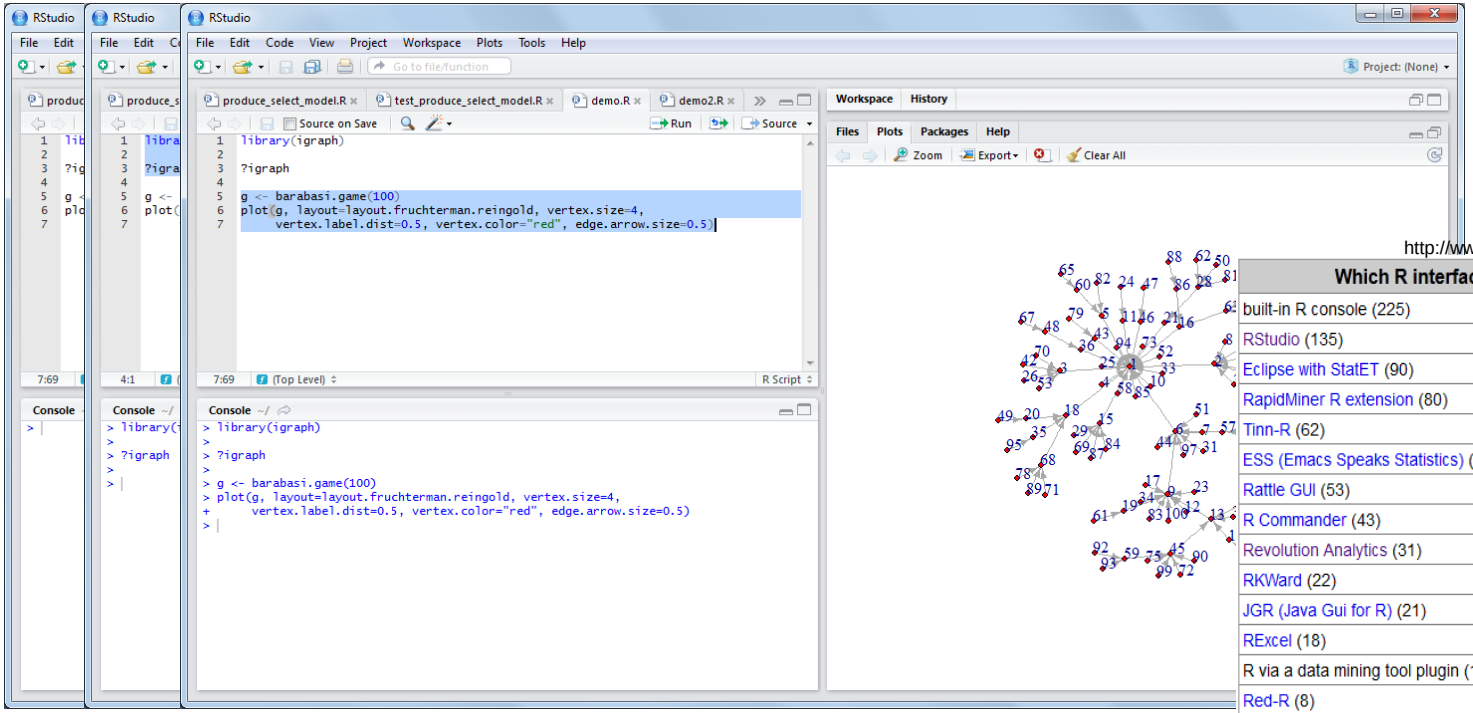
R environment is ..

- Powerful
- Extensible
- Graphical
- Extensive statistics
- OOTB functionality with many ‘knobs’ but smart defaults
- Ease of installation and use
- ***Free***

http://cran.r-project.org/



Third Party Open Source IDEs, e.g., RStudio



http://www.kdnuggets.com/polls/2011/r-gui-used.htm

Which R interfaces do you use frequently?	
built-in R console (225)	40%
RStudio (135)	24%
Eclipse with StatET (90)	16%
RapidMiner R extension (80)	14.2%
Tinn-R (62)	11%
ESS (Emacs Speaks Statistics) (59)	10.5%
Rattle GUI (53)	9.4%
R Commander (43)	7.7%
Revolution Analytics (31)	5.5%
RKWard (22)	3.9%
JGR (Java Gui for R) (21)	3.7%
RExcel (18)	3.2%
R via a data mining tool plugin (12)	2.1%
Red-R (8)	1.4%
SciViews-R (6)	1.1%
Other (44)	7.8%

R's Popularity

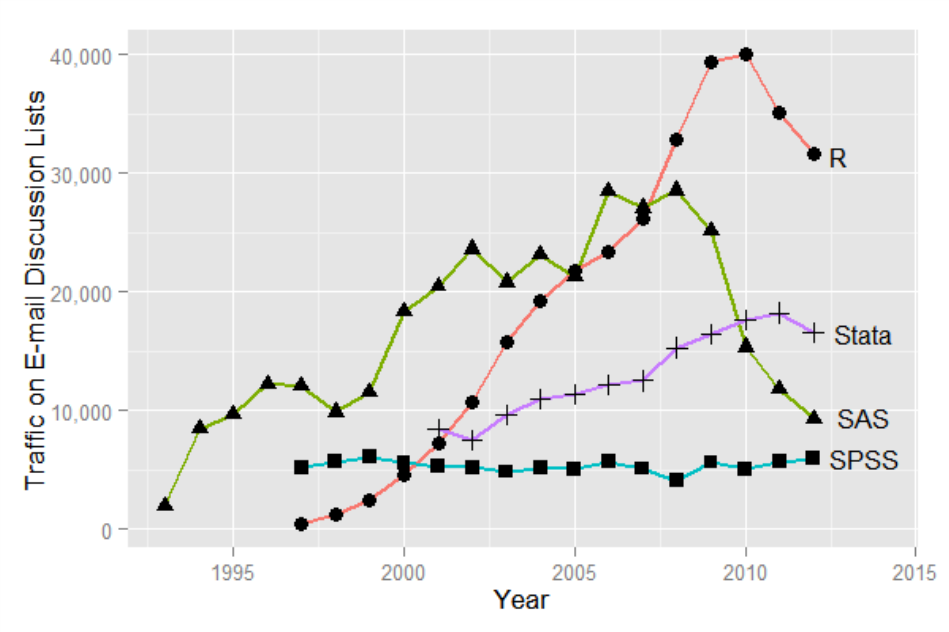


Figure 1a. Sum of monthly email traffic on each software's main listserv discussion list.

"We can see that discussion of R has grown the most rapidly and, for the past few years, R is the most discussed software by an almost two-to-one margin."

http://r4stats.com/articles/popularity/

R's Popularity

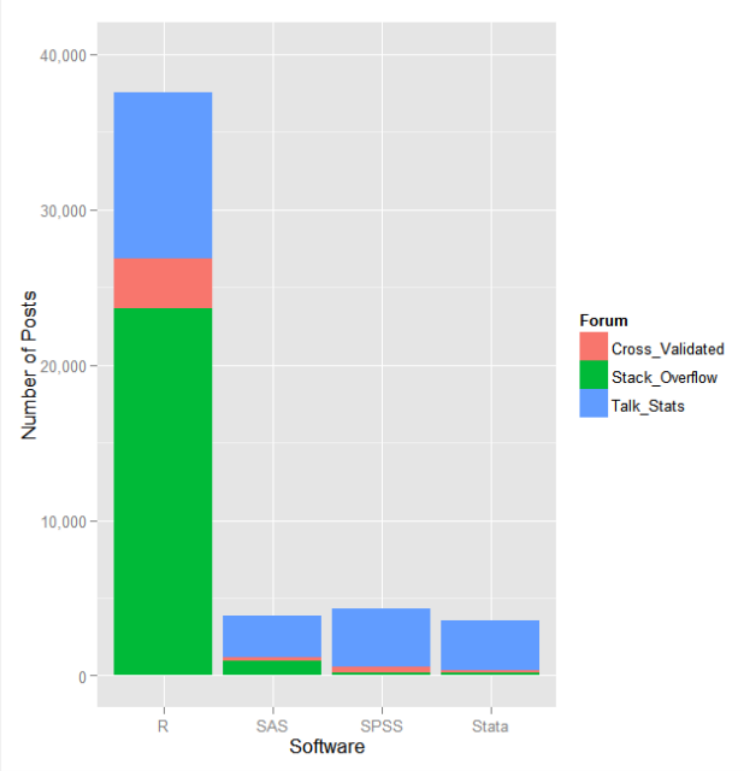


Figure 1b. Number of posts per software on each forum on 2/10/2013.

http://r4stats.com/articles/popularity/

R's Popularity

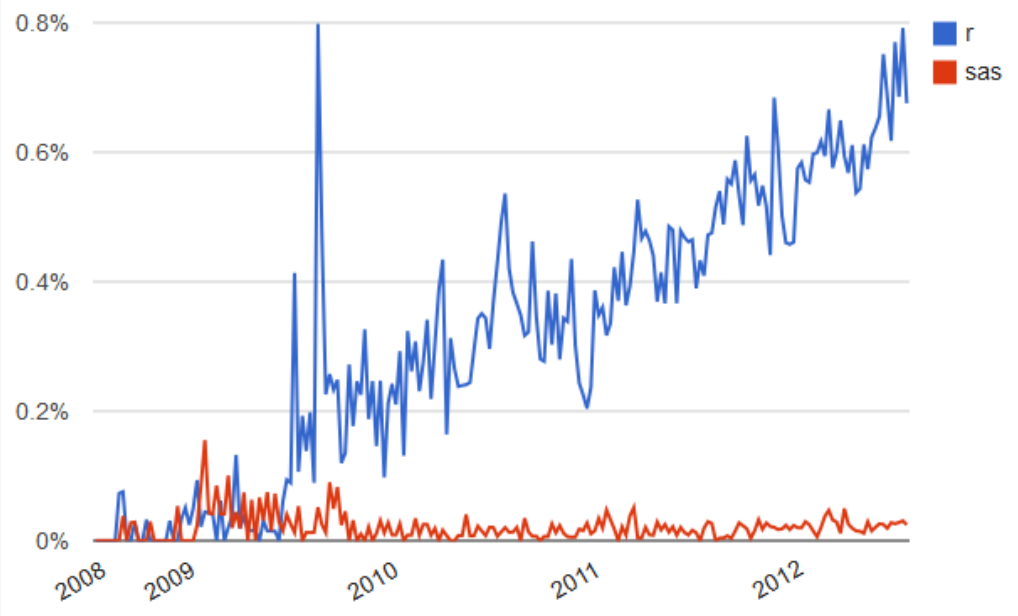
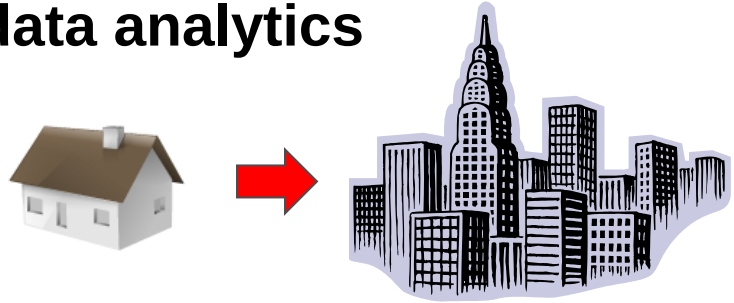


Figure 1c. Number of R- or SAS-related posts to Stack Overflow by week.

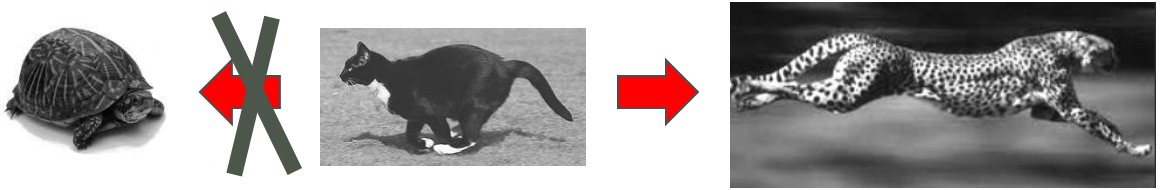
http://r4stats.com/articles/popularity/

Three concerns for enterprise data analytics

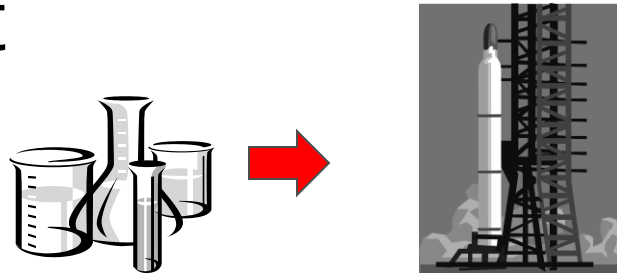
- Scalability



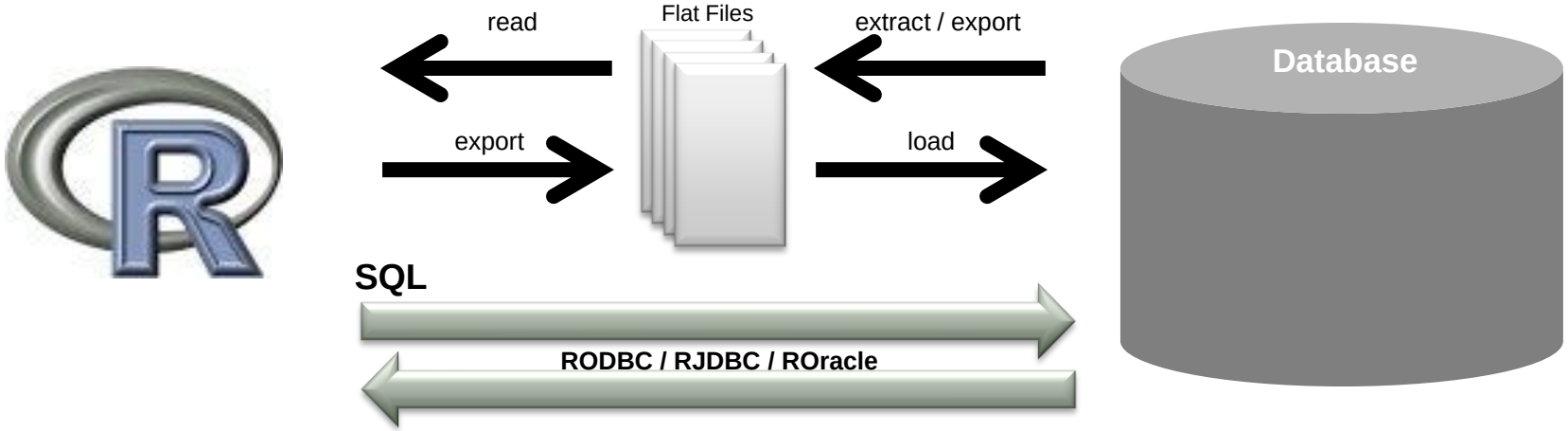
- Performance



- Production Deployment



Traditional R and Database Interaction

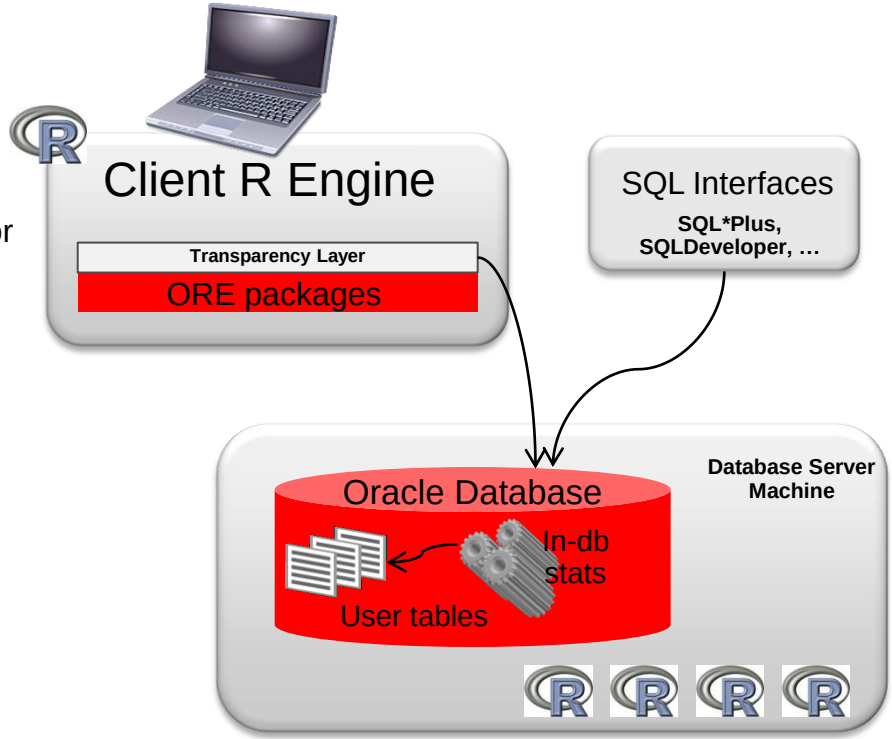


R script
cron job

- Paradigm shift: R → SQL → R
- R memory limitation – data size, call-by-value
- R single threaded
- Access latency, backup, recovery, security…?
- Ad hoc script execution

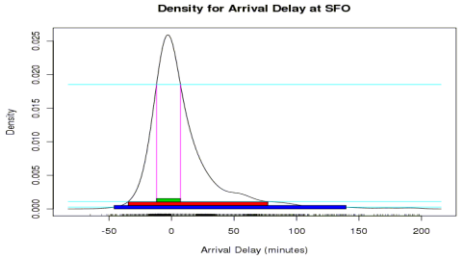
Target Environment with ORE

- A comprehensive, database-centric environment for end-to-end analytical processes in R, with immediate deployment to production environments
- Operationalize entire R scripts in production applications – eliminate porting R code
- Seamlessly leverage Oracle Database as an HPC environment for R scripts, providing data parallelism and resource management
- Avoid reinventing code to integrate R results into existing applications
- Transparently analyze and manipulate data in Oracle Database through R using versatile and customizable R functions
- Eliminate memory constraint of client R engine
- Score R models in Oracle Database
- Execute R scripts through Oracle Database server machine for scalability and performance
- Get maximum value from your Oracle Database and Exadata
- Enable integration and management through SQL
- Integrate R into the IT software stack, e.g. OBIEE

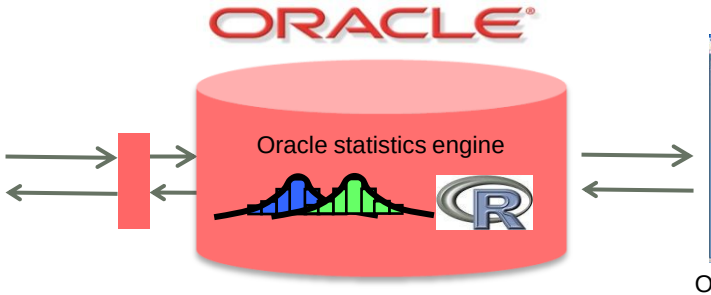


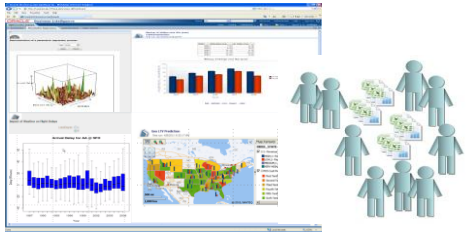
Oracle R Enterprise

R workspace console



Function push-down –
data transformation & statistics





OBIEE, Web Services

No changes to the user experience

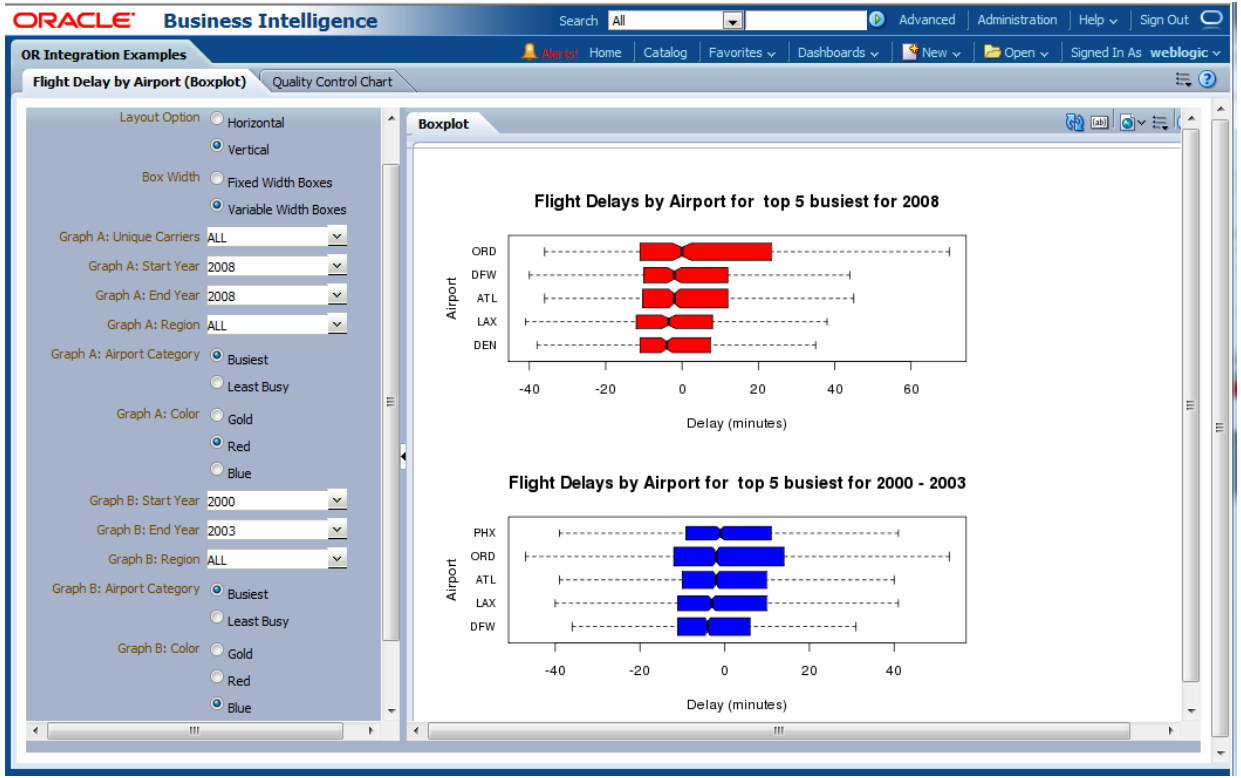
Scale to large data sets

Embed in operational systems

Development → Production → Consumption

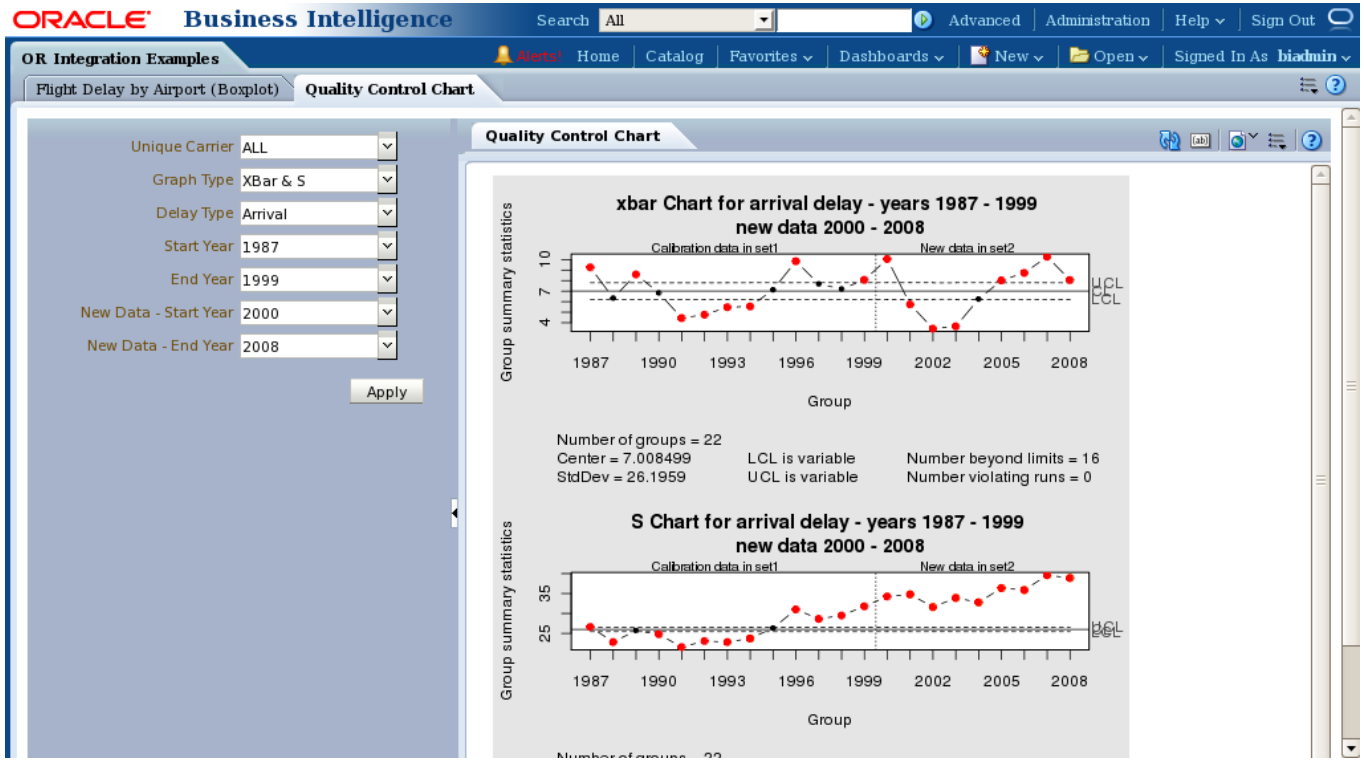
OBIEE Dashboard

Parameterized data selection and graph customization



OBIEE Dashboard

Leverage open source R packages



Oracle's R Strategic Offerings

Deliver enterprise-level advanced analytics based on R environment

- Oracle R Enterprise
  - Transparent access to database-resident data from R
  - Embedded R script execution through database managed R engines with SQL language integration
  - Statistics engine
- Oracle R Distribution
  - Free download, pre-installed on Oracle Big Data Appliance, bundled with Oracle Linux
  - Enterprise support for customers of Oracle R Enterprise, Big Data Appliance, and Oracle Linux
  - Enhanced linear algebra performance using Intel, AMD, or Solaris libraries
- ROracle
  - Open source Oracle *database interface driver* for R based on OCI
  - Maintainer is Oracle – rebuilt from the ground up
  - Optimizations and bug fixes made available to open source community
- Oracle R Advanced Analytics for Hadoop
  - R interface to Oracle Hadoop Cluster on BDA
  - Access and manipulate data in HDFS, database, and file system
  - Write MapReduce functions using R and execute through natural R interface
  - Leverage native MapReduce advanced analytic techniques built on the framework

Oracle R Distribution

Ability to dynamically load

Intel Math Kernel Library (MKL)

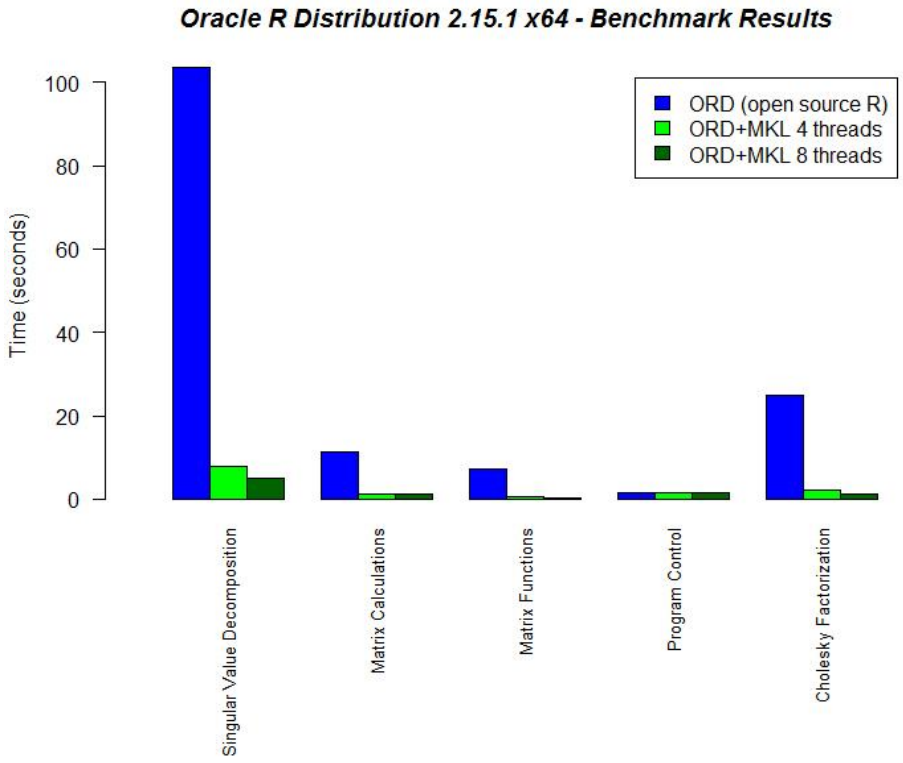
AMD Core Math Library (ACML)

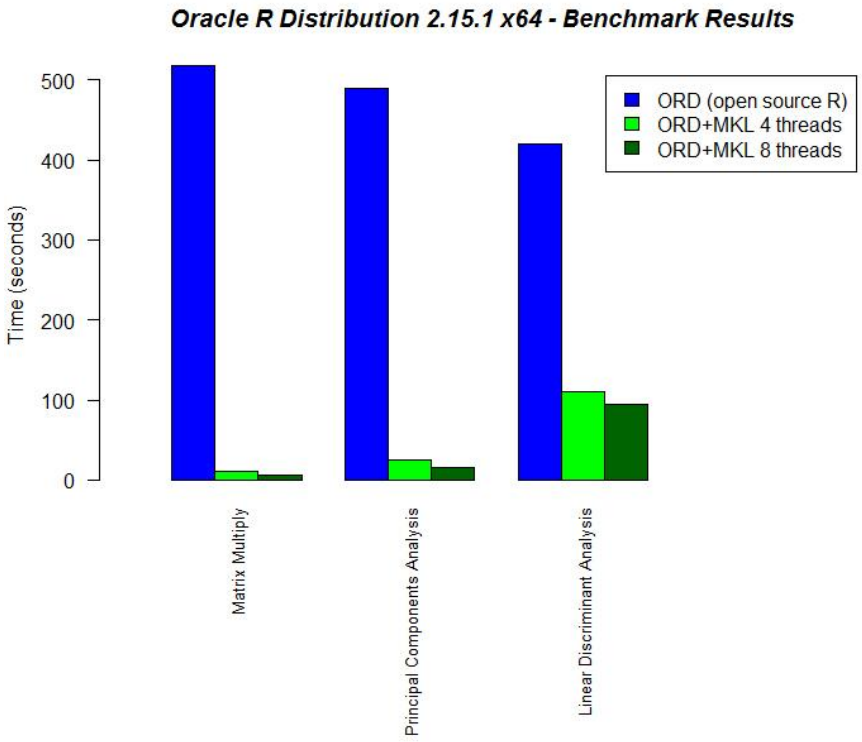
Solaris Sun Performance Library

Oracle Support

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

ORD Performance with MKL

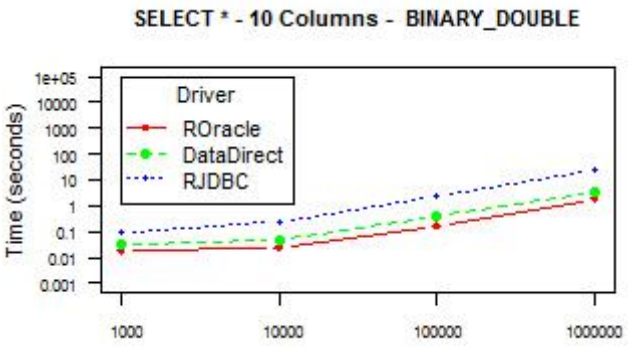


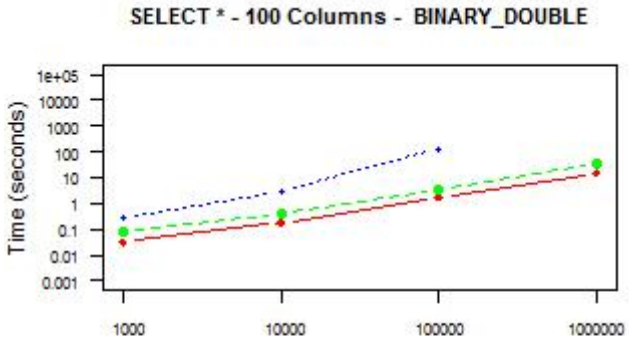


ROracle

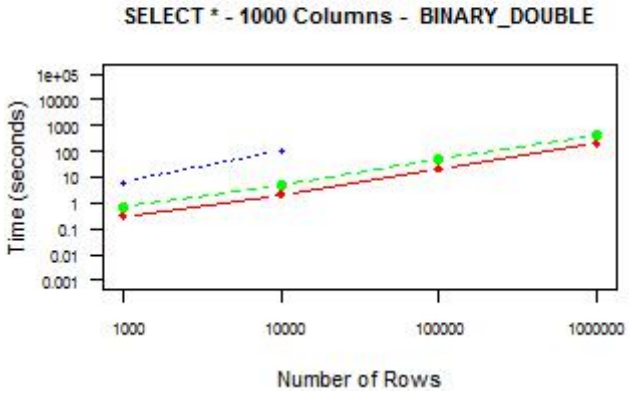
- R package enabling connectivity to Oracle Database
  - Open source
  - Publicly available on CRAN
- Execute SQL statements from R interface
- Enables transactional behavior for insert, update, and delete
- Oracle Database Interface (DBI) for R
- DBI –compliant Oracle driver based on OCI
- Requirements
  - Oracle Instant Client - allows running applications without installing the standard Oracle Database Client or having an ORACLE_HOME. OCI, OCCI, Pro*C, ODBC, and JDBC applications work without modification, while using significantly less disk space. SQL*Plus can also be used with Instant Client.
  - Or, the standard Oracle Database Client

Comparison loading database table to R data.frame





- ROracle
  - Up to 79X faster than RJDBC
  - Up to 2.5X faster than RODBC
  - Scales across NUMBER, VARCHAR2, TIMESTAMP data types



See https://blogs.oracle.com/R/entry/r_to_oracle_database_connectivity

ROracle Example – rolling back transactions

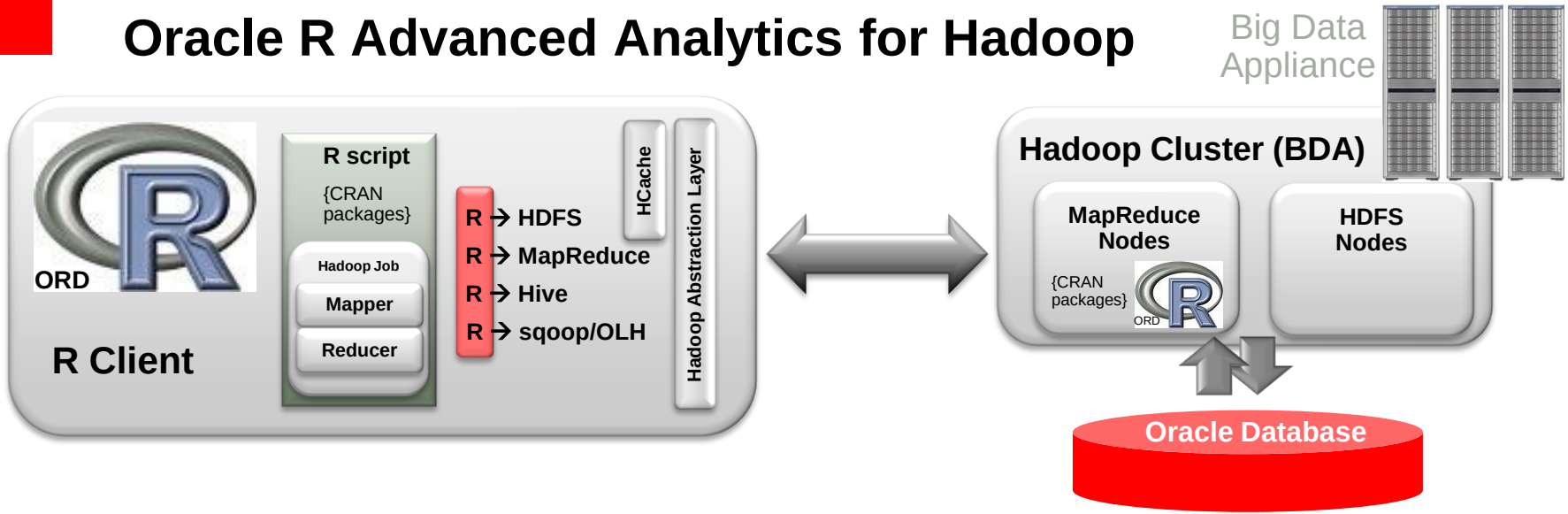
```
drv <- dbDriver("Oracle")
con <- dbConnect(drv, username = "scott", password = "tiger")

dbReadTable(con, "EMP")

rs <- dbSendQuery(con, "delete from emp where deptno = 10")

dbReadTable(con, "EMP")
if(dbGetInfo(rs, what = "rowsAffected") > 1){
   warning("dubious deletion -- rolling back transaction")
   dbRollback(con)
}
dbReadTable(con, "EMP")
```

Oracle R Advanced Analytics for Hadoop

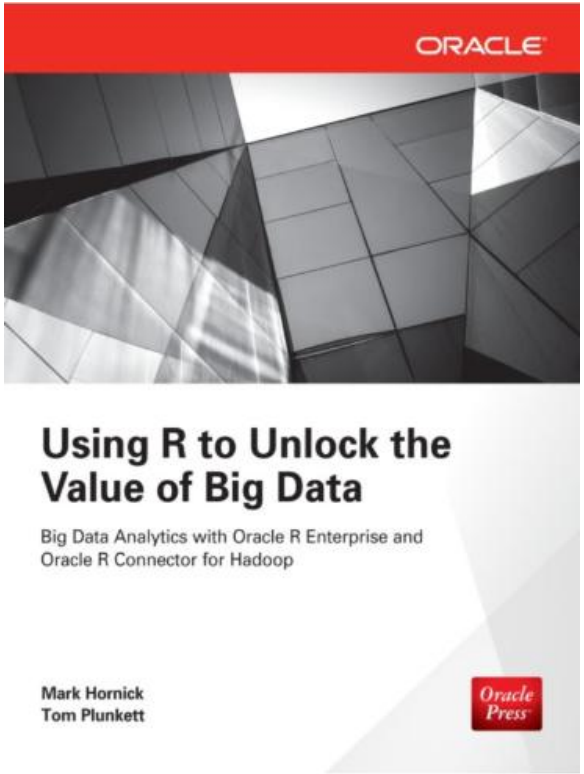


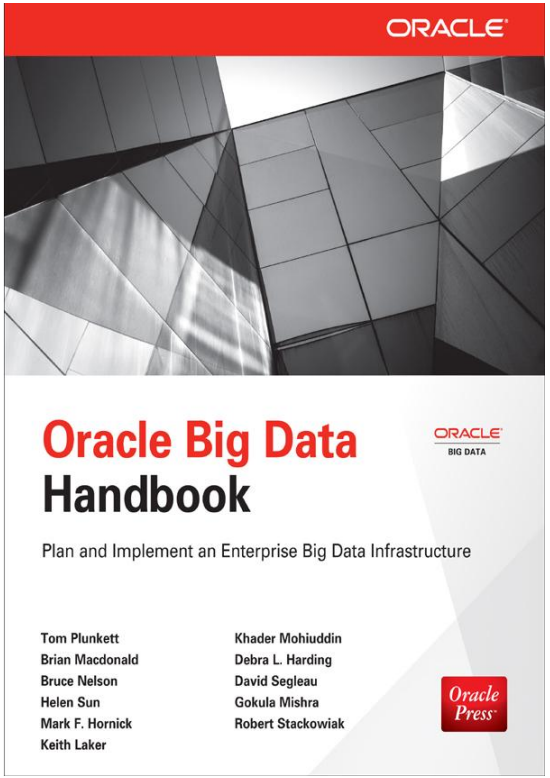
- Expand user population that can build models on Hadoop
- Accelerate rate at which business problems are tackled
- Deliver analytics that scale with data volumes, variables, techniques

- Provide transparent access to Hadoop Cluster
- Manipulate data in HDFS, database, and file system - all from R
- Write and execute MapReduce jobs with R
- Leverage CRAN R packages to work on HDFS-resident data
- Move from lab to production without requiring knowledge of Hadoop internals, Hadoop CLI, or IT infrastructure

ORAAH Analytics Functions

Function	Description
orch.cor	Correlation matrix computation
orch.cov	Covariance matrix computation
orch.kmeans	Perform k-means clustering on a data matrix stored as an HDFS file. Score data using orch.predict.
orch.lm	Fits a linear model using tall-and-skinny QR (TSQR) factorization and parallel distribution. The function computes the same statistical parameters as the Oracle R Enterprise ore.lm function. Score data using orch.predict.
orch.lmf	Fits a low rank matrix factorization model using either the jellyfish algorithm or the Mahout alternating least squares with weighted regularization (ALS-WR) algorithm.
orch.neural	Provides a neural network to model complex, nonlinear relationships between inputs and outputs, or to find patterns in the data. Score data using orch.predict.
orch.nmf	Provides the main entry point to create a nonnegative matrix factorization model using the jellyfish algorithm. This function can work on much larger data sets than the R NMF package, because the input does not need to fit into memory.
orch.princomp	Principal components analysis of HDFS data. Score data using orch.predict.
orch.sample	Sample HDFS data by percentage or explicit number of rows specification





Oracle R Enterprise – Brief Introduction

Transparency Layer

Aggregation function on ore.frame object

```
aggdata <- aggregate(ONTIME_S$DEST,
                     by = list(ONTIME_S$DEST),
                     FUN = length)
class(aggdata)
head(aggdata)
```

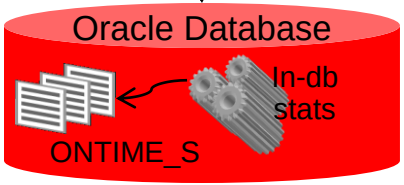
```
R> aggdata <- aggregate(ONTIME_S$DEST,
+                       by = list(ONTIME_S$DEST),
+                       FUN = length)
R> class(aggdata)
[1] "ore.frame"
attr(,"package")
[1] "OREbase"
R> head(aggdata)
  Group.1    x
0     ABE  237
1     ABI   34
2     ABQ 1357
3     ABY   10
4     ACK    3
5     ACT   33
```

Client R Engine

Transparency Layer

Oracle R package

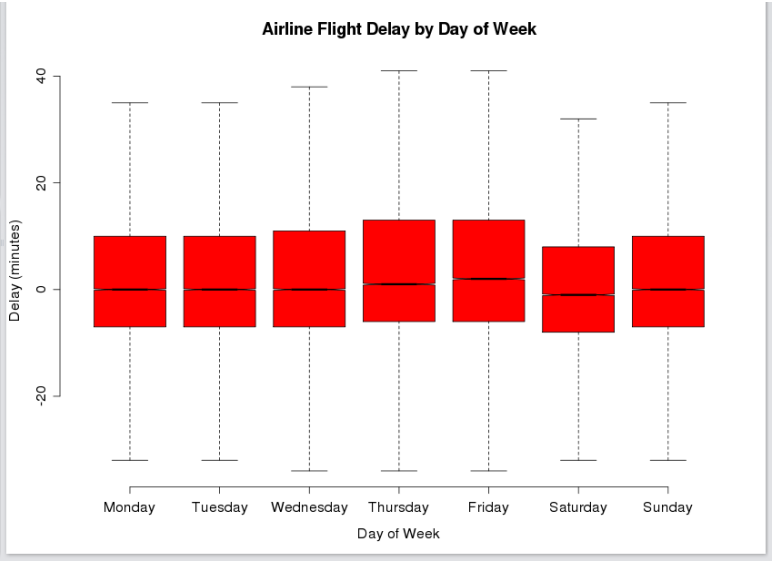
```
select DEST, count(*)
from ONTIME_S
group by DEST
```



Transparency Layer

Overloads graphics functions for in-database statistics

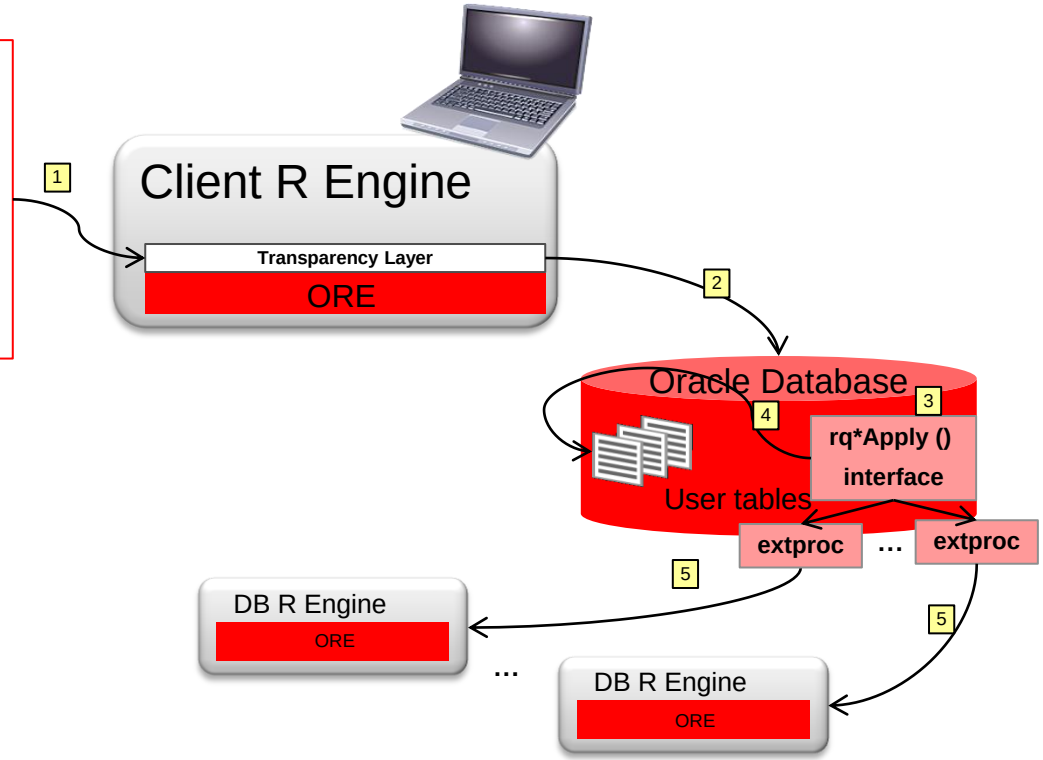
```
ontime <- ONTIME_S
delay <- ontime$ARRDELAY
dayofweek <- ontime$DAYOFWEEK
bd <- split(delay, dayofweek)
boxplot(bd, notch = TRUE, col = "red", cex = 0.5,
        outline = FALSE, axes = FALSE,
        main = "Airline Flight Delay by Day of Week",
        ylab = "Delay (minutes)", xlab = "Day of Week")
axis(1, at=1:7, labels=c("Monday", "Tuesday",
                         "Wednesday", "Thursday",
                         "Friday", "Saturday", "Sunday"))
axis(2)
```



Embedded R Execution – R Interface

Data parallel in-database execution

```
modList <- ore.groupApply(
   X=ONTIME_S,
   INDEX=ONTIME_S$DEST,
   function(dat) {
      lm(ARRDELAY ~ DISTANCE + DEPDELAY, dat)
   });
modList_local <- ore.pull(modList)
summary(modList_local$BOS) ## return model for BOS
```



Also includes

- ore.doEval
- ore.tableApply
- ore.rowApply
- ore.indexApply

Embedded R Execution – SQL Interface

For model build and batch scoring

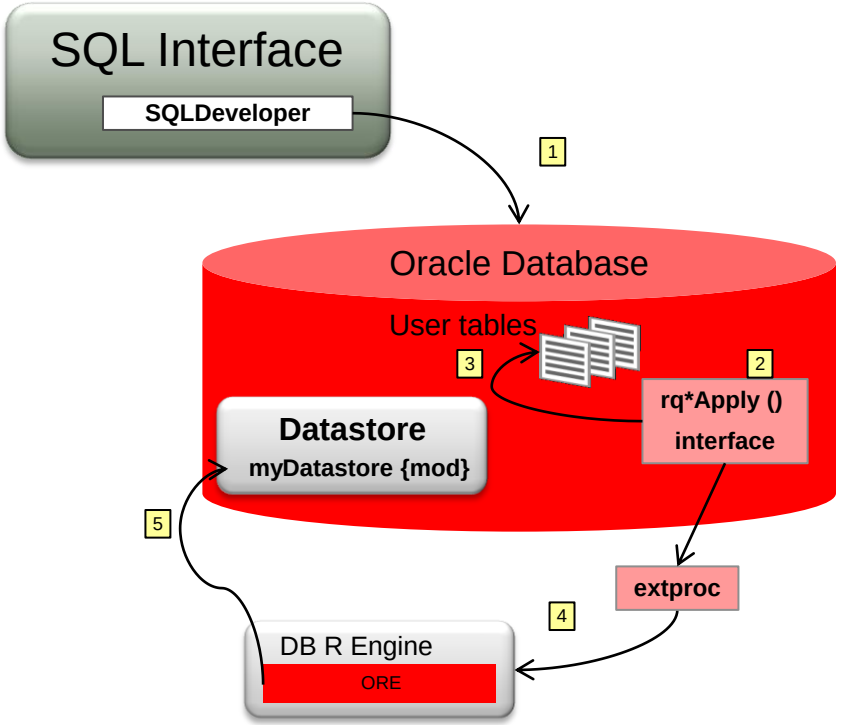
```
begin
  sys.rqScriptDrop('Example2');
  sys.rqScriptCreate('Example2',
 'function(dat,datastore_name) {
   mod <- lm(ARRDELAY ~ DISTANCE + DEPDELAY, dat)
   ore.delete(datastore_name)
   ore.save(mod,name=datastore_name)
  }');
end;
/

select *
  from table(rqTableEval(
     cursor(select ARRDELAY,
                   DISTANCE,
                   DEPDELAY
            from   ontime_s),
     cursor(select 1 "ore.connect",
                   'myDatastore' as "datastore_name"
            from dual),
     'XML',
     'Example2' ));
```

```
begin
  sys.rqScriptCreate('Example3',
 'function(dat, datastore_name) {
     ore.load(datastore_name)
     prd <- predict(mod, newdata=dat)
     prd[as.integer(rownames(prd))] <- prd
     res <- cbind(dat, PRED = prd)
     res}');
end;
/
select *
from table(rqTableEval(
    cursor(select ARRDELAY, DISTANCE, DEPDELAY
           from   ontime_s
           where  year = 2003
           and    month = 5
           and    dayofmonth = 2),
    cursor(select 1 "ore.connect",
                  'myDatastore' as "datastore_name" from dual),
    'select ARRDELAY, DISTANCE, DEPDELAY, 1 PRED from ontime_s',
    'Example3'))
order by 1, 2, 3;
```

Embedded R Execution – SQL Interface

rqTableEval + datastore for model building



Statistics Engine

Example Features

- 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

Oracle R Enterprise 1.4 Key New Features

- ORE-specific algorithms for high performance and scalability
  - Neural Networks: **ore.neural** highly flexible network architecture with wide range of activation functions
  - Generalized Linear Models: **ore.glm**
  - Exponential Smoothing (simple and double): **ore.esm**
  - ore.neural, ore.glm, ore.stepwise and ore.lm leverage embedded R parallelism, exceeding 1000 column formula derived column limit, and supporting categorical variables
  - Support for weights in regression models
  - Factor Analysis - **factanal** Transparency Layer function
  - Principal Component Analysis - **princomp** Transparency Layer function
  - ANOVA - **anova** Transparency Layer function on ore.lm fit object
  - Support for discretization of numeric variables using **cut** Transparency Layer function
- OREdm package
  - **ore.odmAssocRules** - Association Rules with igraph package compatibility
  - **ore.odmNMF** - Non-Negative Matrix Factorization
  - **ore.odmOC** – O-Cluster

Oracle R Enterprise 1.4 Key New Features (2)

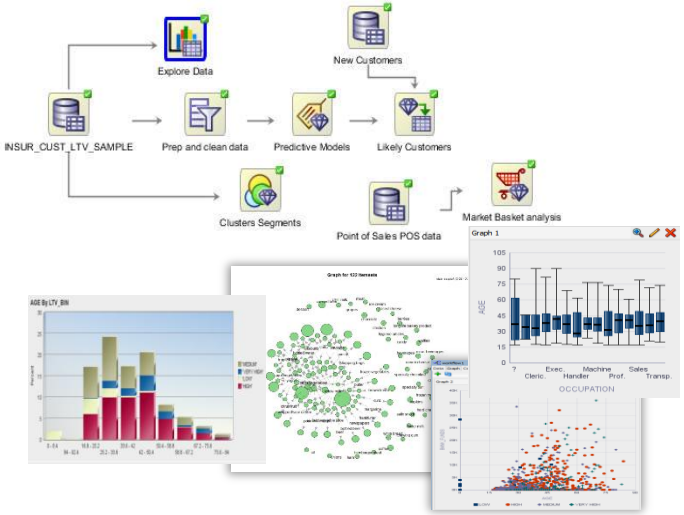
- Embedded R Execution enhancements
  - Optimized interface to load data from several sources into Oracle database
  - Support for efficient discretization of numeric variables
  - explicitly specify degree of parallelism (DOP) for parallel-enabled functions (ore.groupApply, ore.rowApply, ore.indexApply), use ore.options(ore.parallel=n)
- Embedded R graphics rendering performance enhancement
  - cairo device - R graphics rendered in image buffer, eliminating temp file writes
  - eliminates need for X11 configuration for graphics on Unix-like ORE server side
- Support for migration of ORE R script repository and datastore across databases
  - Supports ease of production deployment from development environments
  - Supports snapshotting of production environments for evaluation/debugging in test systems
- Improved performance for table creation from R - ore.create()
  - Enables execution of CTAS in parallel
  - Always executes with NOLOGGING
  - Leverages ore.parallel R option to specify parallelism

Oracle Advanced Analytics Option

Oracle Advanced Analytics Option

Fastest Way to Deliver Scalable Enterprise-wide Predictive Analytics

- Powerful
  - Accelerate rate at which business problems are tackled
  - Improve time to insight
  - Combination of in-database predictive algorithms and open source R algorithms
  - Accessible via SQL, PL/SQL, R and database APIs
  - Scalable, parallel in-database execution of R language
- Easy to Use
  - Expand user population that can build models
  - Range of GUI and IDE options for business users to data scientists
- Enterprise-wide
  - Integrated feature of Oracle Database via SQL - R is integrated into SQL
  - Seamless support for enterprise analytical applications / BI environments

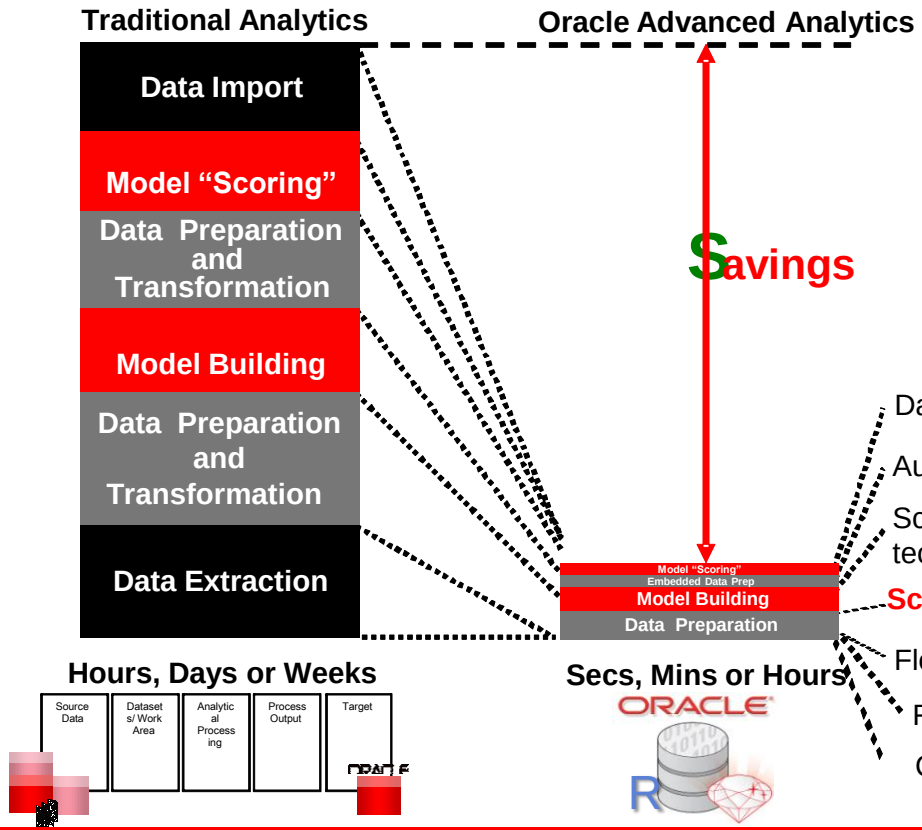


Oracle R Enterprise

+

Oracle Data Mining

Oracle Advanced Analytics Value Proposition



Value Proposition

- Fastest path from data to insights
- Fastest analytical development
- **Fastest in-database scoring engine**
- **Flexible deployment options for analytics**
- Lowest TCO by eliminating data duplication
- Secure, Scalable and Manageable

Data remains in the Database

Automated data preparation for select analytics

Scalable distributed-parallel implementation of machine learning techniques in-database

Scalable R leveraging database computational engine

Flexible interface options – R, SQL, IDE, GUI

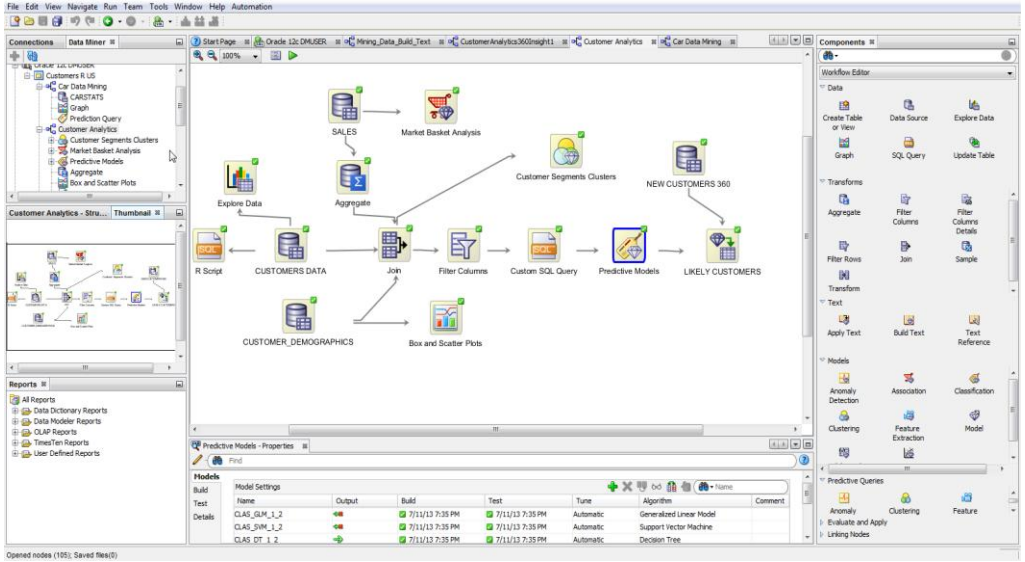
Fastest and most Flexible analytic deployment options

Can import 3rd party models

Oracle Data Miner GUI

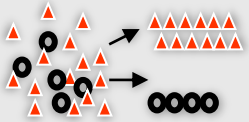
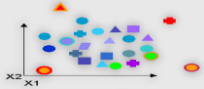
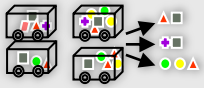
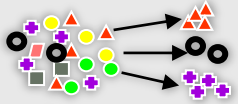
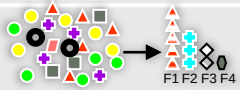
SQL Developer 4.0 Extension—*Free OTN Download*

- Easy to Use
  - Oracle Data Miner GUI for data analysts
  - “Work flow” paradigm
- Powerful
  - Multiple algorithms & data transformations
  - Runs 100% in-DB
  - Build, evaluate and apply models
- Automate and Deploy
  - Generate SQL scripts for deployment
  - Share analytical workflows

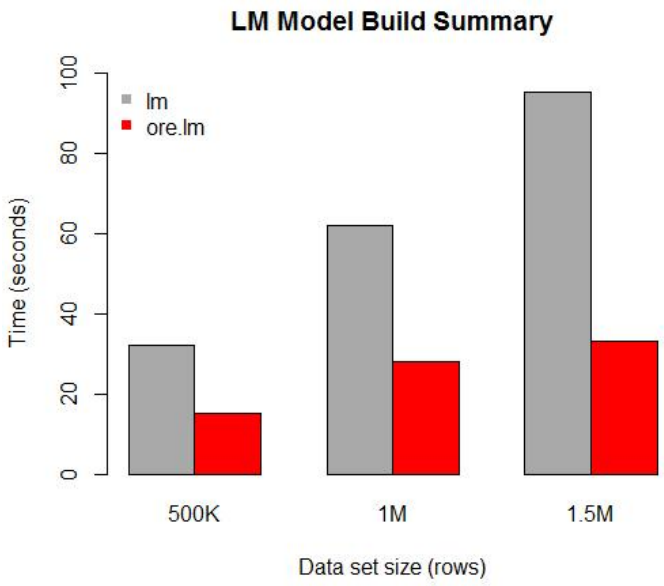


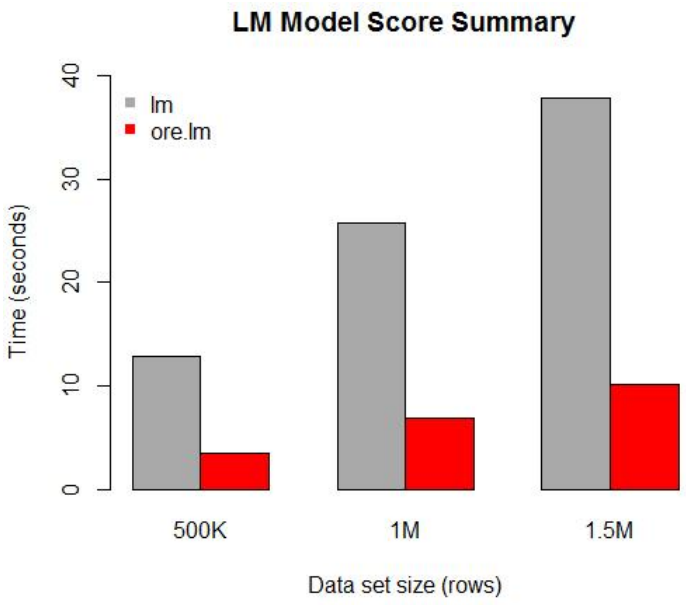
Oracle Advanced Analytics

In-Database Data Mining Algorithms from Oracle Data Mining

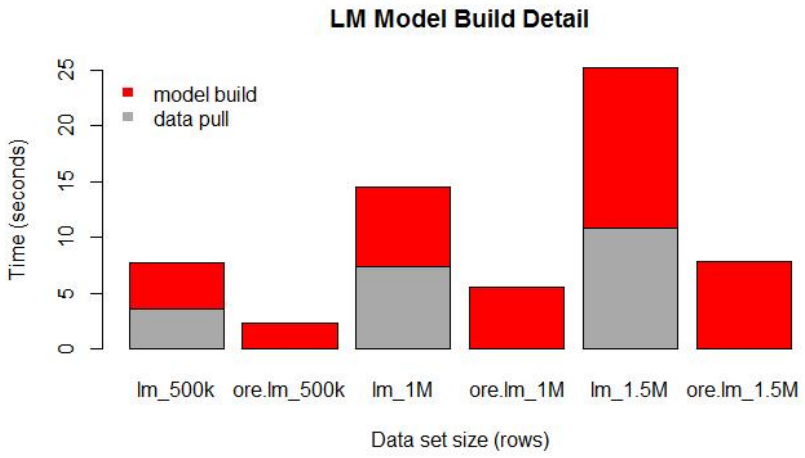
		Algorithms	Applicability
Classification		Logistic Regression (GLM) Decision Trees Naïve Bayes Support Vector Machines (SVM)	Classical statistical technique Popular / Rules / transparency Embedded app Wide / narrow data / text
Regression		Linear Regression (GLM) Support Vector Machine (SVM)	Classical statistical technique Wide / narrow data / text
Anomaly Detection		One Class SVM	Unknown fraud cases or anomalies
Attribute Importance		Minimum Description Length (MDL) Principal Components Analysis (PCA)	Attribute reduction, Reduce data noise
Association Rules		Apriori	Market basket analysis / Next Best Offer
Clustering		Hierarchical k-Means Hierarchical O-Cluster Expectation-Maximization Clustering (EM)	Product grouping / Text mining Gene and protein analysis
Feature Extraction		Nonnegative Matrix Factorization (NMF) Singular Value Decomposition (SVD)	Text analysis / Feature reduction

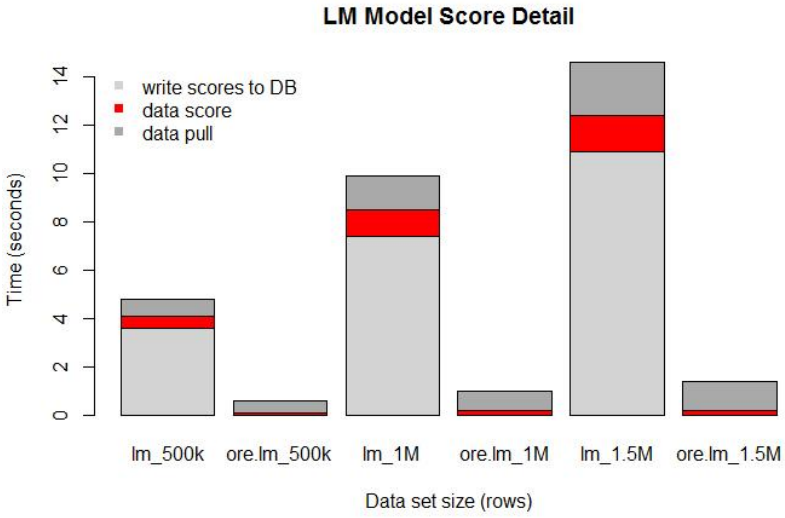
In-database Performance Advantage



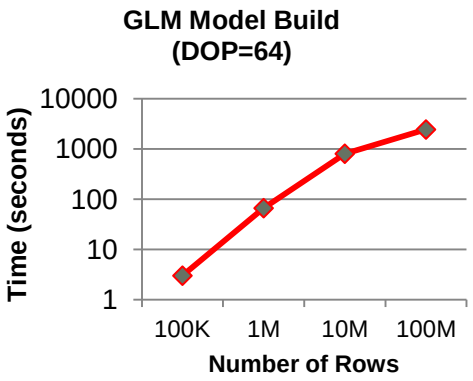


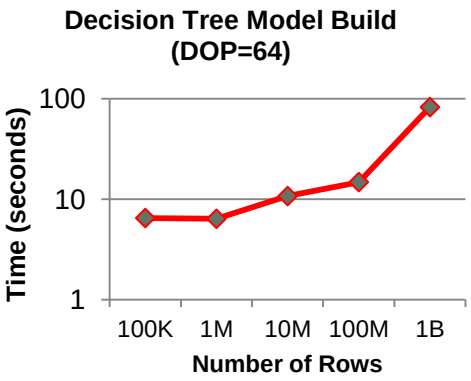
In-database Performance Advantage

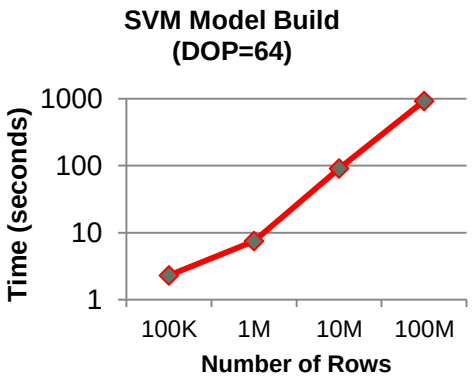




In-database GLM, Decision Tree, and SVM







11 seconds to score 100M records (DOP=64)

92 seconds to score 1B records (DOP=128)

13 seconds to score 1B records (DOP=128)

2466 seconds to score 1B records (DOP=64)

Exadata X2-2 half-rack

12c Parallel Distributed Advanced Analytics

Real world proof points

- Linear Regression (`ore.lm`) on Exadata X3-2 half-rack
  - Data set: 2.9 billion rows spanning 12 months of data with over 350 predictors
  - Elapsed time ~5 minutes!
- Logistic Regression (`ore.glm`) on Exadata X3-2 half-rack
  - Data set: 2.9 billion rows spanning 12 months of data with over 350 predictors
  - Elapsed time ~30 minutes!
- Neural networks (`ore.neural`) on T5-4 Solaris
  - Data set: 1 billion rows with 40 columns
  - Elapsed time ~6 minutes with 10 hidden neurons & 421 weights

Oracle R Enterprise as framework for Advanced Analytics

Workflow example

Analysis

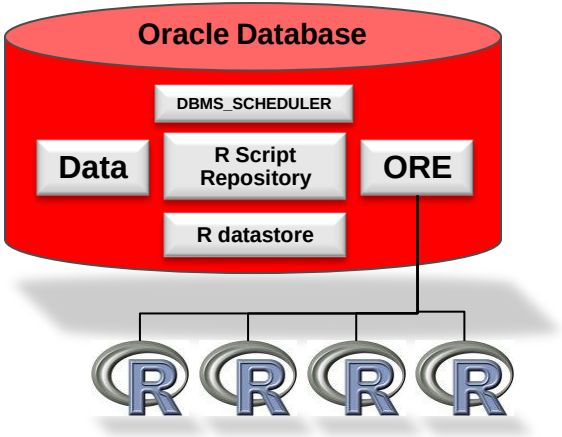
Exploratory Data Analysis, Visualization and Data Preparation → Sample data and split in train and test → Build and test models in parallel with ORE Embedded R Execution → Select best model and save in database 'datastore' object → Load and test model from datastore for scoring new data

Development

Code the build methodology in R script repository → Code the scoring methodology in R script repository → Invoke build and scoring R functions using ore.*Apply

Production

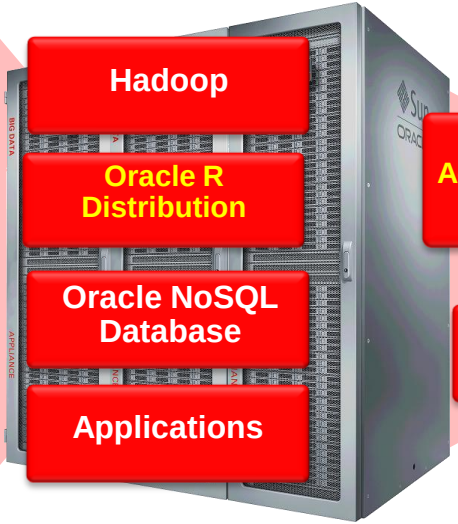
Deploy scripts and R objects from Lab to Production → Schedule build and score as nightly jobs for execution



Oracle Big Data Platform

Oracle Big Data Appliance

Optimized for Hadoop, R, and NoSQL Processing



- Hadoop
- Oracle R Distribution
- Oracle NoSQL Database
- Applications

Oracle Big Data Connectors

- Oracle R Advanced Analytics for Hadoop + …
- Oracle Data Integrator

Oracle Exadata

"System of Record" Optimized for DW/OLTP

- Oracle R Enterprise
- Oracle Data Mining
- Data Warehouse
- Oracle Database
- Oracle R Distribution

Oracle Exalytics

Optimized for Analytics & In-Memory Workloads

- Oracle Enterprise Performance Management
- Oracle Business Intelligence Applications
- Oracle Business Intelligence Tools
- Oracle Endeca Information Discovery

Stream | Acquire | Organize | Discover & Analyze

Summary

- Oracle enables R users with advanced analytics on Big Data
  - Via Oracle Database with Oracle Advanced Analytics – Oracle R Enterprise
  - Via Big Data Appliance with Oracle R Advanced Analytics for Hadoop
- Oracle's R technologies extend R for Enterprise use
  - Data analysis and exploration
  - Application development
  - Production deployment
- Enabling high performance, scalability, and ease of production deployment

Resources

- **Book:** Using R to Unlock the Value of Big Data, by Mark Hornick and Tom Plunkett
- **Blog:** https://blogs.oracle.com/R/
- **Forum:** https://forums.oracle.com/forums/forum.jspa?forumID=1397
- **Oracle R Distribution**
- **ROracle**
- **Oracle R Enterprise**
- **Oracle R Advanced Analytics for Hadoop**

http://oracle.com/goto/R

ORACLE®