

Published on *CERN openlab* (<http://test-static-05.web.cern.ch>)

[Home](#) > Massive Predictive Modeling using Oracle R Enterprise

Massive Predictive Modeling using Oracle R Enterprise ^[1]

Date:

Tuesday, 8 April, 2014 - 11:00 to 12:30

Location:

[CERN IT Auditorium](#) ^[2]

R is fast becoming the lingua franca for analyzing data via statistics, visualization, and predictive analytics. For enterprise-scale data, R users have three main concerns: scalability, performance, and production deployment. Oracle's R-based technologies - Oracle R Distribution, Oracle R Enterprise, Oracle R Connector for Hadoop, and the R package ROracle - address these concerns.

In this talk, we introduce Oracle's R technologies, highlighting how each enables R users to achieve scalability and performance while making production deployment of R results a natural outcome of the data analyst/scientist efforts. The focus then turns to Oracle R Enterprise with code examples using the transparency layer and embedded R execution, targeting massive predictive modeling. One goal behind massive predictive modeling is to build models per entity, such as customers, zip codes, simulations, in an effort to understand behavior and tailor predictions at the entity level. Predictions can then be aggregated, for example, to assess future demand. Massive predictive modeling comes with challenges: effectively partitioning data, where to store and manage the resulting models, how to associate models with customers, as well as backup, recovery, and security.

While R has parallel capabilities to facilitate taking advantage of clusters of computers, significant coding is usually required to meet the challenges noted above. In this talk, we present the business problem and illustrate how Oracle R Enterprise, one of Oracle's R technologies, facilitates massive predictive modeling in a pair of succinct R scripts. With Oracle R Enterprise, the data, R scripts, and models all reside in Oracle Database.

About the speaker

Mark Hornick, Director, Oracle Advanced Analytics, focuses on Oracle's R Technologies. He works with internal and external customers in the application of R for scalable advanced

analytics applications in Oracle Database, Exadata, and the Big Data Appliance. Mark is coauthor of the books *Using R to Unlock the Value of Big Data* and *Oracle Big Data Handbook*, published by Oracle Press. He joined Oracle's Data Mining Technologies group in 1999 through the acquisition of Thinking Machines Corp. Mark also evangelizes and conducts training sessions on Oracle's R technologies internationally, and has presented at conferences including Oracle OpenWorld, Collaborate, BIWA Summit, and user!. Mark holds a Bachelor's degree from Rutgers University and a Master's degree from Brown University, both in Computer Science.

Indico or other event webpage:

[For more information \(slides and video available\)](#) ^[3]

- [Visit Us](#)
- [RSS Feeds](#)

DISCLAIMER: This Web page contains pointers to material related to the management of CERN openlab in the Information Technology Department at the European Organization for Nuclear Research (CERN). Their use and distribution are regulated by the [CERN copyright notice](#).



Source URL: <http://test-static-05.web.cern.ch/news/massive-predictive-modeling-using-oracle-r-enterprise>

Links

[1] <http://test-static-05.web.cern.ch/news/massive-predictive-modeling-using-oracle-r-enterprise>

[2] <http://maps.cern.ch/mapsearch/?centerX=2492493¢erY=1121078¢erScale=1500>

[3] <http://indico.cern.ch/event/309959/>