

使用R机器学习打造智能零售行业的创新体验

董乃文 Nevin Dong

资深技术专家

微软开发工具及平台事业部(DX)

nevind@microsoft.com

议题

- 零售业的挑战及数字化转型
- 微软云高级分析服务及 R 平台
- Hadoop + Spark + R：三驾齐驱
- R + 机器学习：助力最佳零售数字化体验

零售业及品牌面临挑战，需要提供全面的线上线下最佳数字化体验



人口结构变化



传统作业系统



直达消费者



全周期顾客管理

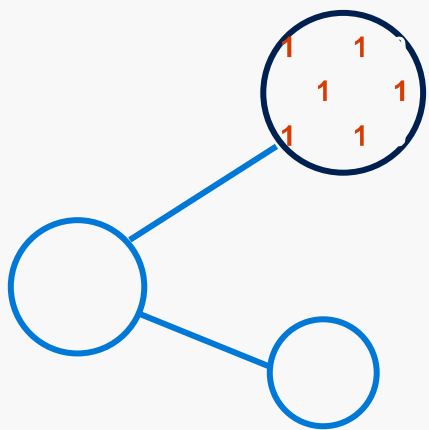


信息化消费者



数字化体验

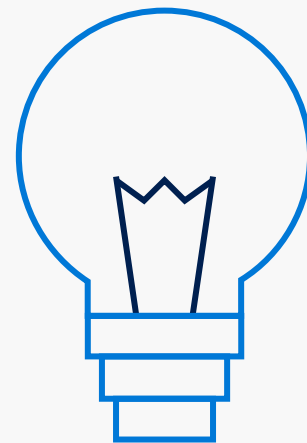
业务数字化转型趋势



大数据



云计算



人工智能

消费者体验驱动的现代购物环境

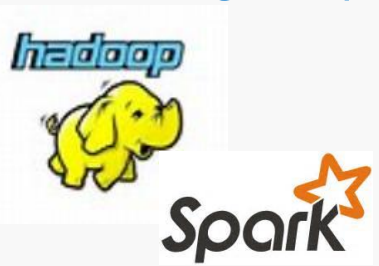


微软云高级分析服务及 R 平台

微软云高级分析服务解决方案

解决方案

HDInsight/Spark



大数据平台

大规模并行计算 企业级规模，一次写入，
分布式部署

数据科学家/
数据工程师

Microsoft R



基于R的分析

数据科学家

Azure 机器学习



云分析

便于建模，分析
及对比

数据科学家

认知服务

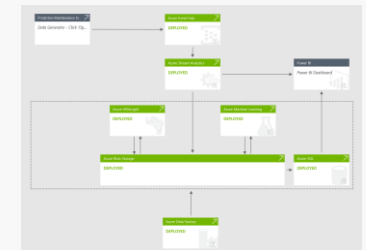


分析API服务

便捷的视觉、语音、语言、
知识等的分析

开发者

解决方案



预配置的应用及
解决方案

解决特定行业/业务/领域
的应用/解决方案

BDM/TDM

Microsoft R 兵器谱

- Microsoft R Open
- Microsoft R Client
- Microsoft R Server...
...for Hadoop, ...for Teradata, ...for Linux (SUSE, Red Hat/CentOS)
- Microsoft SQL Server 2016 R Services
- Microsoft R Open in Azure ML

至尊魔戒 Microsoft R: One Ring to Rule Them All

Forbes / Tech

JAN 15, 2016 @ 06:14 AM 7,803 VIEWS

Microsoft R: One Big Data Tool To Rule Them All?

 **Adrian Bridgwater**
CONTRIBUTOR

I track enterprise software application development & data management.

[FOLLOW ON FORBES \(72\)](#)

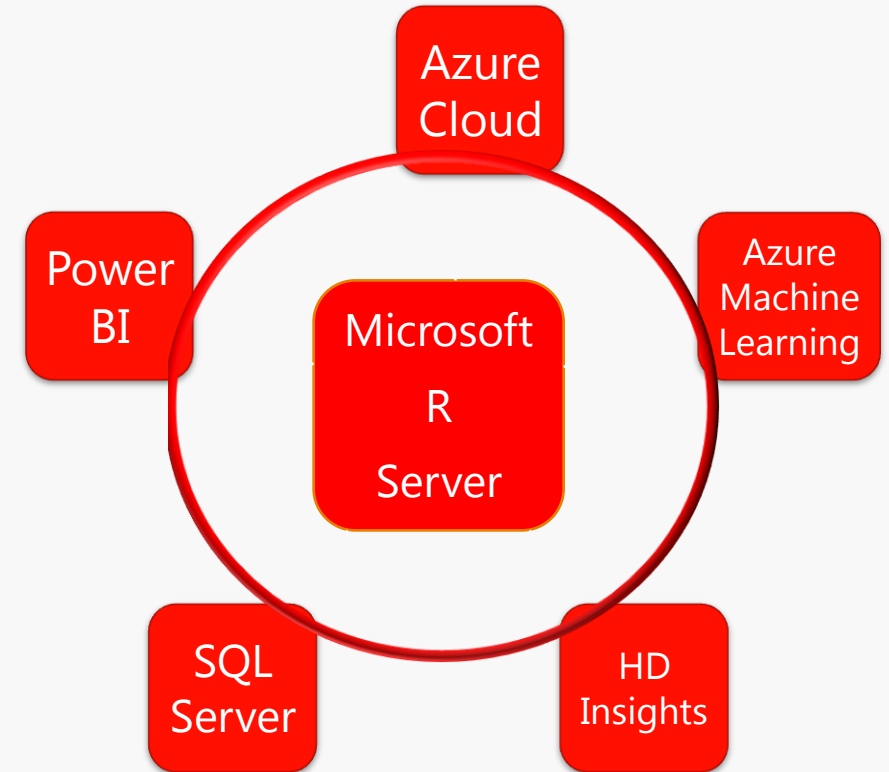
[FULL BIO >](#)

Opinions expressed by Forbes Contributors are their own.



'Microsoft R Open' — a product name almost worth getting T-shirts printed for, were it not grammatically incorrect. Redmond's big data analytics dream builds one tool to rule them all, maybe... Image: Wikipedia

Microsoft MSFT +0.75% wants a slice of the big data analytics pie. Truth be told, it has already baked and served itself up a portion by acquiring the R-language and data crunching specialist Revolution Analytics, a purchase it completed in spring of 2015.

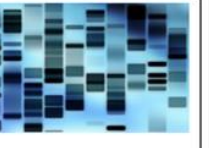
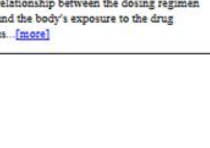
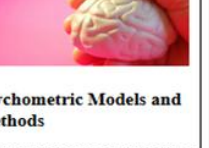


R + CRAN

(8,000+ packages)

CRAN Task Views

CRAN Task Views are guides to the packages and functions useful for certain disciplines and methodologies. Many long-term R users I know have no idea they exist. As an effort to make them more widely known I thought I'd jazz up the index page. Images are free to use, and got from [SXC](#) stock photo site. Visual puns are mine. Task View links go to the cran.r-project.org site and not a mirror.

 Bayesian Inference Applied researchers interested in Bayesian statistics are increasingly attracted to R because of the ease of which one can code algorithms to sample. [more]	 Chemometrics and Computational Physics Chemometrics and computational physics are concerned with the analysis of data arising in chemistry and physics experiments, as well as the simulation of. [more]	 Clinical Trial Design, Monitoring, and Analysis This task view gathers information on specific R packages for design, monitoring and analysis of data from clinical trials. It focuses on including. [more]	 Cluster Analysis & Finite Mixture Models This CRAN Task View contains a list of packages that can be used for finding groups in data and modelling unobserved cross-sectional heterogeneity. Many... [more]	 Probability Distributions For most of the classical distributions, base R provides probability distribution functions (p), density functions (d), quantile functions (q), and. [more]	 Computational Econometrics Base R ships with a lot of functionality useful for computational econometrics, in particular in the stats package. This functionality is complemented by many... [more]	 Analysis of Ecological and Environmental Data This Task View contains information about using R to analyse ecological and environmental data. [more]	 Design of Experiments (DoE) & Analysis of Experimental Data This task view collects information on R packages for experimental design and analysis of data from experiments. Please feel free to suggest enhancements. [more]	 Empirical Finance This CRAN Task View contains a list of packages useful for empirical work in Finance, grouped by topic. [more]	 Statistical Genetics Great advances have been made in the field of genetic analysis over the last years. The availability of millions of single nucleotide polymorphisms (SNPs). [more]
 Natural Language Processing This CRAN task view contains a list of packages useful for natural language processing. [more]	 Analysis of Pharmacokinetic Data The primary goal of pharmacokinetic (PK) data analysis is to determine the relationship between the dosing regimen and the body's exposure to the drug as. [more]	 Official Statistics & Survey Methodology This CRAN task view contains a list of packages that includes methods typically used in official statistics and survey methodology. Many packages provide... [more]	 Phylogenetics, Especially Comparative Methods The history of life unfolds within a phylogenetic context. Comparative phylogenetic methods are statistical approaches for analyzing historical. [more]	 Multivariate Statistics Base R contains most of the functionality for classical multivariate analysis, somewhere. There are a large number of packages on CRAN which extend this... [more]	 Optimization and Mathematical Programming This CRAN task view contains a list of packages which offer facilities for solving optimization problems. Although every regression model in statistics. [more]	 Machine Learning & Statistical Learning Several add-on packages implement ideas and methods developed at the borderline between computer science and statistics - this field of research is usually. [more]	 Graphic Displays & Dynamic Graphics & Graphic Devices & Visualization R is rich with facilities for creating and developing interesting graphics. Base R contains functionality for many plot types including coplots, mosaic. [more]	 High-Performance and Parallel Computing with R This CRAN task view contains a list of packages, grouped by topic, that are useful for high-performance computing (HPC) with R. In this context, we are. [more]	 Medical Image Analysis This task view is for input, output, and analysis of medical imaging files. [more]
 Analysis of Spatial Data Base R includes many functions that can be used for reading, visualising, and analysing spatial data. The focus in this view is on "geographical" spatial. [more]	 Survival Analysis Survival analysis, also called event history analysis in social science, or reliability analysis in engineering, deals with time until occurrence of an. [more]	 Time Series Analysis Base R ships with a lot of functionality useful for time series, in particular in the stats package. This is complemented by many packages on CRAN, which are... [more]	 Robust Statistical Methods Robust (or "resistant") methods for statistics modelling have been available in S from the start, in R in package stats (e.g., median(), mean(), trim =). [more]	 Statistics for the Social Sciences Social scientists use a wide range of statistical methods. To make the burden carried by this task view lighter, I have suppressed detail in some areas that... [more]	 gRaphical Models in R Wikipedia defines a graphical model as a graph that represents independencies among random variables by a graph in which each node is a random variable, and. [more]	 Reproducible Research The goal of reproducible research is to tie specific instructions to data analysis and experimental data so that scholarship can be recreated, better. [more]	 Psychometric Models and Methods Psychometrics is concerned with the design and analysis of research and the measurement of human characteristics. Psychometricians have also worked... [more]		

Microsoft R Server

R+CRAN

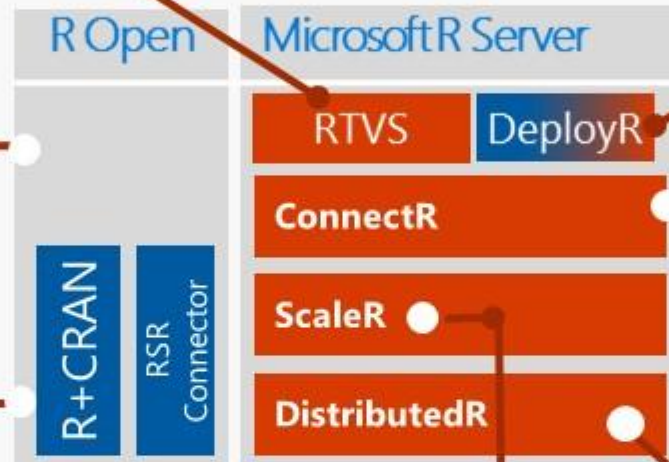
- Open source R interpreter
 - R 3.1.2
- Freely-available huge range of R algorithms
- Algorithms callable by RevoR
- Embeddable in R scripts
- 100% Compatible with existing R scripts, functions and packages

Microsoft R Open

- Based on open source R
- High-performance math library to speed up linear algebra functions
- Checkpoint package to easily share R code and replicate results using specific R package versions

R Tools for Visual Studio

- State of the art, R Tools for Visual Studio IDE



DeployR

- RESTful APIs for easy integration from Java, JavaScript, .NET
- Enterprise authentication & security
- Horizontal scaling

ConnectR

- High-speed & direct connectors

Available for:

- High-performance XDF
- SAS, SPSS, delimited & fixed format text data files
- Hadoop HDFS (text & XDF)
- Teradata Database & Aster
- EDWs and ADWs
- ODBC

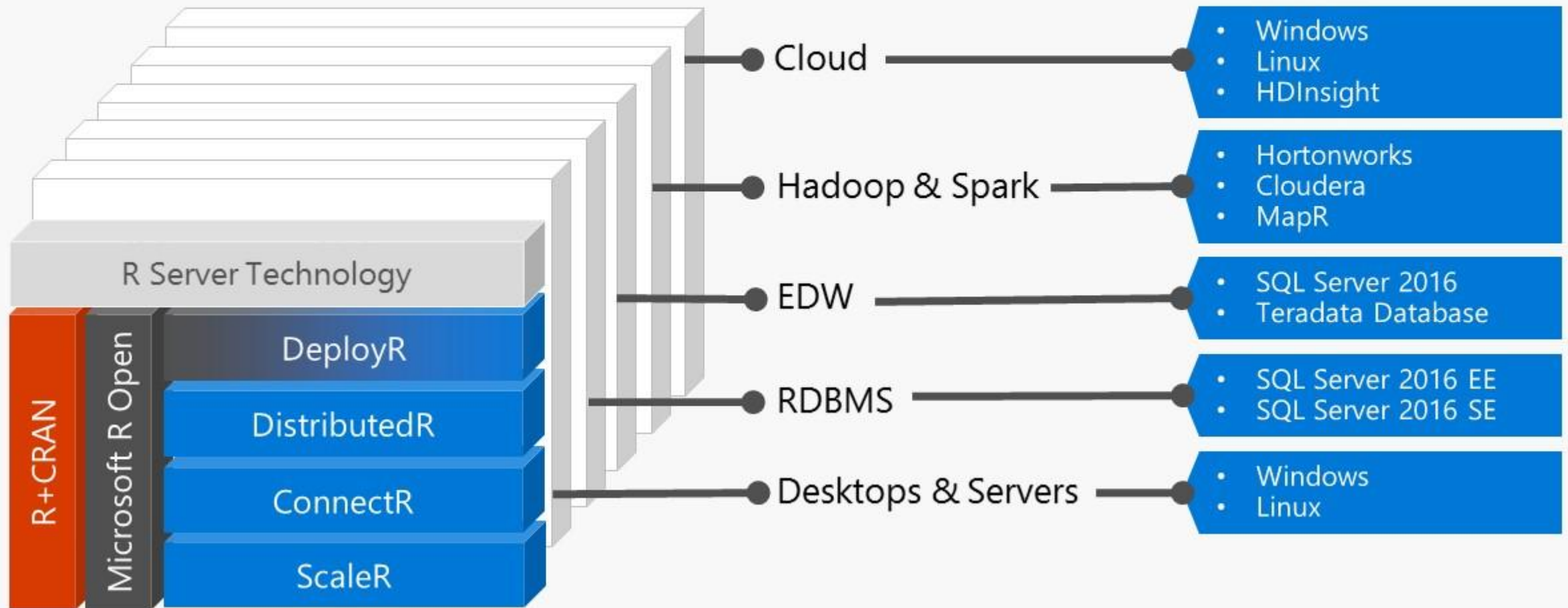
ScaleR

- Ready-to-Use high-performance big data big analytics
- Fully-parallelized analytics
- Data prep & data distillation
- Descriptive statistics & statistical tests
- Range of predictive functions
- User tools for distributing customized R algorithms across nodes
- Wide data sets supported – thousands of variables

DistributedR

- Distributed computing framework
- Delivers cross-platform portability

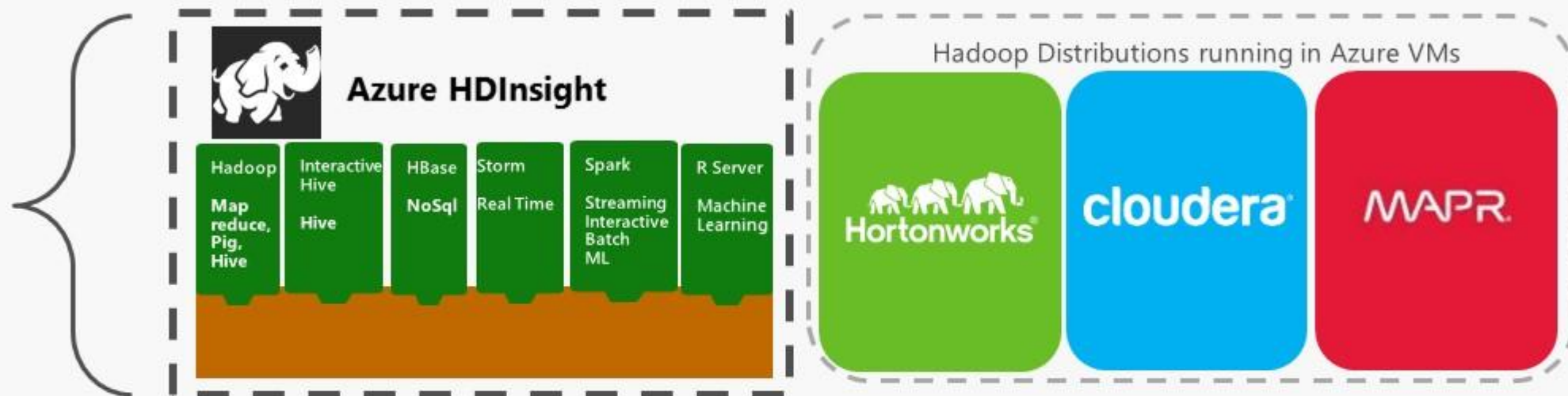
Microsoft R Server



Hadoop + Spark + R : 三驾齐驱

Microsoft Hadoop 技术栈

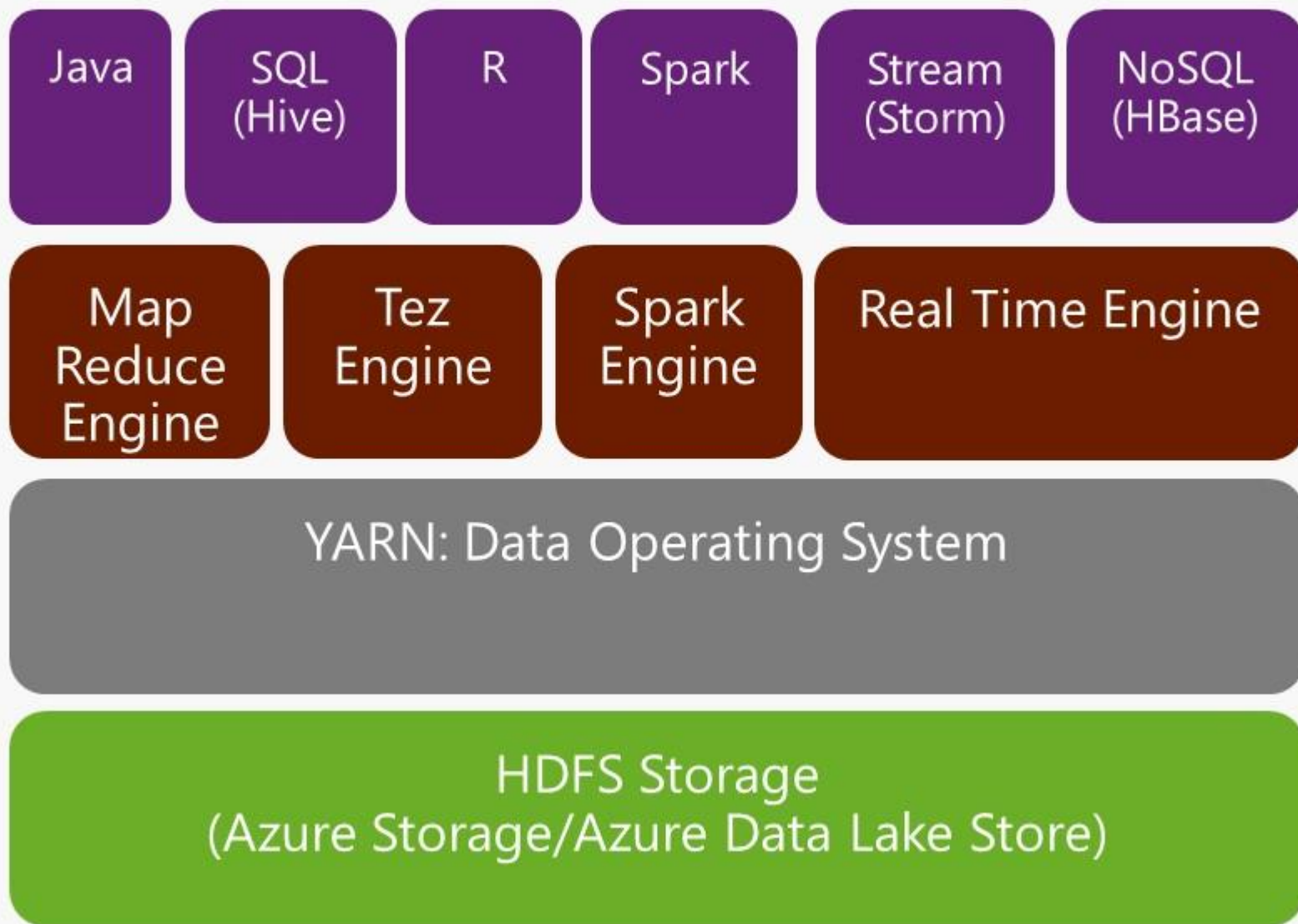
分析



存储

Local (HDFS) or Cloud (Azure Blob/Azure Data Lake Store)

HDInsight 开放的技术架构



基于HDI Insight提供高伸缩性的机器学习平台



Cluster configuration

* Cluster Type ⓘ

R Server

* Operating System

Linux

* Version

R Server 9.0 (HDI 3.5)

* Cluster Tier ⓘ

STANDARD PREMIUM

R Server : Terabyte-scale, enterprise grade R analytics with transparent parallelization on top of Spark and Hadoop.

Configuration Options:

- R Server 9.0.0 on Spark 2.0.0 with Java 8
- R Server 8.0.5 on Spark 1.6.2 with Java 7

Adds \$0.02 per Core-Hour.

Features

* denotes preview feature

Available	Not available
☒ R Studio community edition for R Server	+ Apache Ranger* (PREMIUM) ⓘ
+ Secure shell (SSH) access	+ Domain joining* (PREMIUM) ⓘ
+ HDInsight applications	+ Remote Desktop access ⓘ
+ Custom virtual network	
+ Custom Hive metastore	
+ Custom Oozie metastore	
+ Data Lake Store access	



Cluster configuration

* Cluster Type ⓘ

Spark

* Operating System

Linux

* Version

Spark 2.0.1 (HDI 3.5)

* Cluster Tier ⓘ

STANDARD PREMIUM

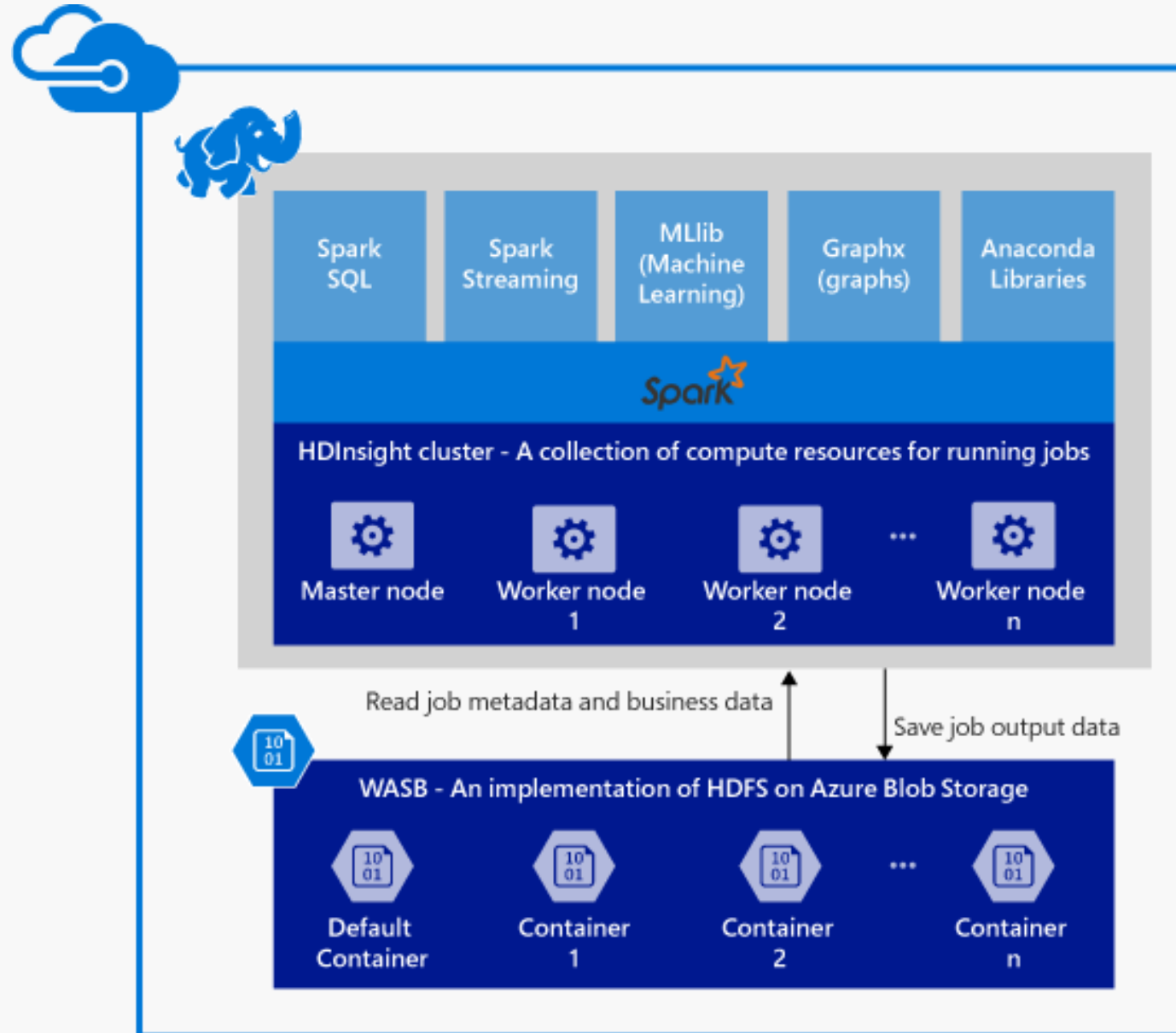
Spark : Fast data analytics and cluster computing using in-memory processing.

Features

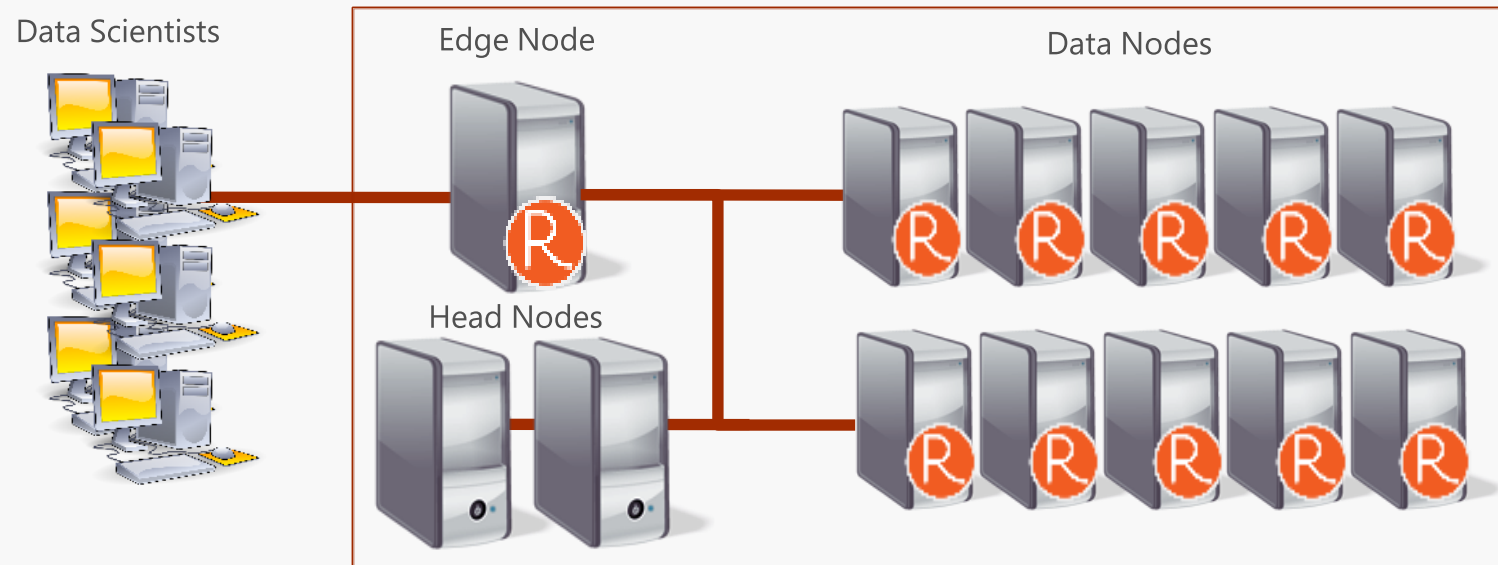
* denotes preview feature

Available	Not available
+ Secure shell (SSH) access	+ Apache Ranger* (PREMIUM) ⓘ
+ HDInsight applications	+ Domain joining* (PREMIUM) ⓘ
+ Custom virtual network	+ Remote Desktop access ⓘ
+ Custom Hive metastore	
+ Custom Oozie metastore	
+ Data Lake Store access	

Spark clusters in Azure HDInsight

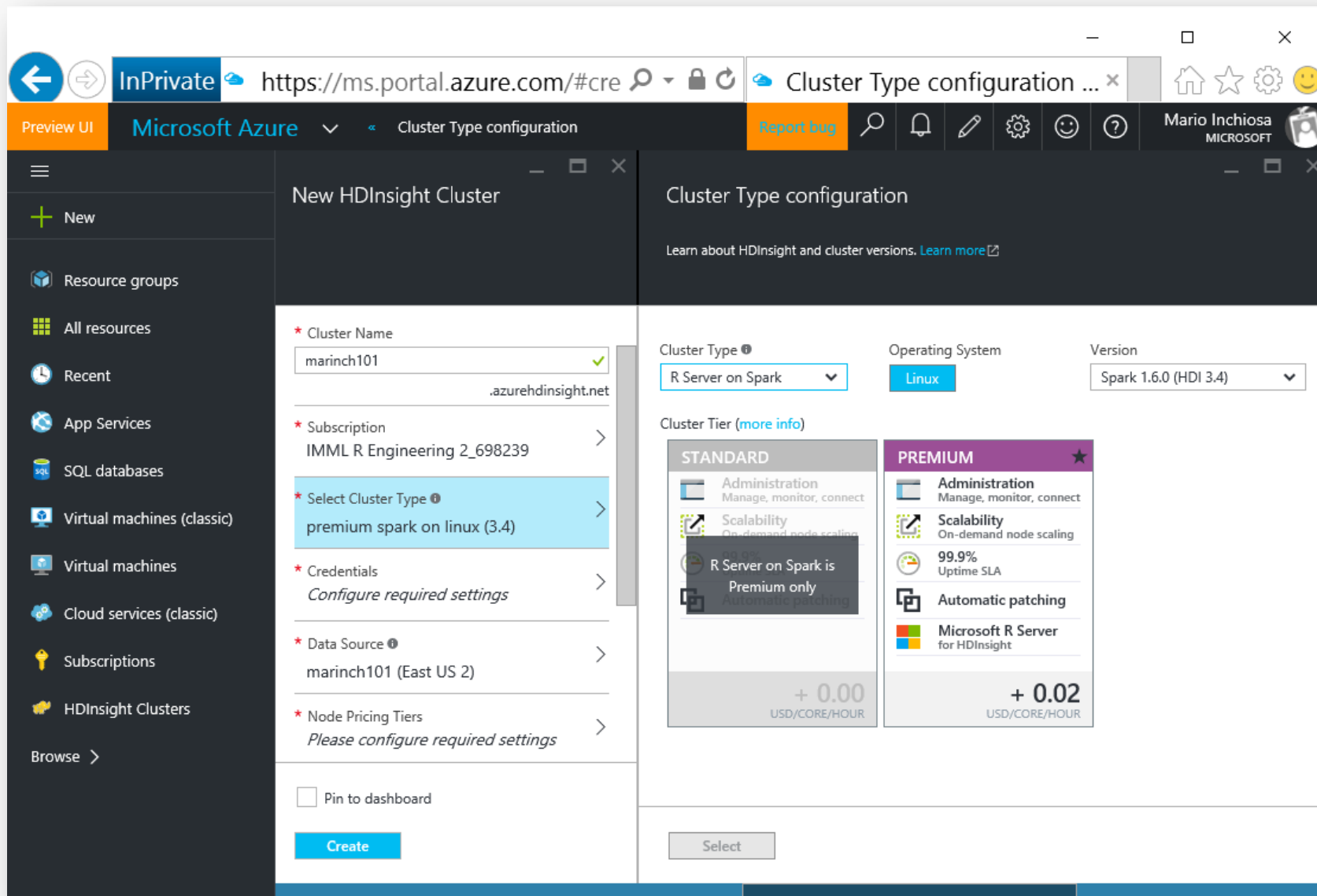


R Server on HDInsight



R Server

创建 HDInsight cluster - R Server 类型



The screenshot displays the Azure portal interface for creating a new HDInsight cluster. The left sidebar shows the navigation menu with options like 'New', 'Resource groups', 'All resources', 'Recent', 'App Services', 'SQL databases', 'Virtual machines (classic)', 'Virtual machines', 'Cloud services (classic)', 'Subscriptions', and 'HDInsight Clusters'. The main content area is titled 'New HDInsight Cluster' and includes a 'Cluster Type configuration' section. The configuration details are as follows:

- Cluster Name:** marinch101
- Subscription:** IMML R Engineering 2_698239
- Select Cluster Type:** premium spark on linux (3.4)
- Credentials:** Configure required settings
- Data Source:** marinch101 (East US 2)
- Node Pricing Tiers:** Please configure required settings

The 'Cluster Type configuration' section shows the following settings:

- Cluster Type:** R Server on Spark
- Operating System:** Linux
- Version:** Spark 1.6.0 (HDI 3.4)
- Cluster Tier:** PREMIUM (selected over STANDARD)

The 'PREMIUM' tier is highlighted with a star and shows the following features:

- Administration:** Manage, monitor, connect
- Scalability:** On-demand node scaling
- 99.9% Uptime SLA**
- Automatic patching**
- Microsoft R Server for HDInsight**

The pricing for the PREMIUM tier is shown as + 0.02 USD/CORE/HOUR. The STANDARD tier is shown as + 0.00 USD/CORE/HOUR. A 'Select' button is visible at the bottom of the configuration section.

集群扩展

The screenshot shows the 'Scale Cluster' settings page in the Microsoft Azure portal. The page is divided into three main sections: a left sidebar with navigation links, a central settings pane, and a right-hand configuration pane.

Left Sidebar: Contains a search bar, a list of settings categories (SUPPORT & TROUBLESHOOTING, GETTING STARTED, CONFIGURATION), and a 'Scale Cluster' tile.

Central Settings Pane: Displays a search bar and a list of settings categories. The 'Scale Cluster' option is highlighted.

Right-hand Configuration Pane: Shows the current configuration for the cluster. The 'Number of Worker nodes' is set to 4. The 'Worker Nodes Pricing Tier' is D4 (4 nodes, 32 cores). The 'Head Node Pricing Tier' is D4 (2 nodes, 16 cores). A cost estimation table is displayed below.

Category	Configuration	Cost (USD/HOUR)
WORKER NODES	1.24 x 4 = 4.97	
HEAD NODES	1.24 x 2 = 2.49	
TOTAL COST		7.46

Cost Summary:

- TOTAL COST:** 7.46 USD/HOUR (ESTIMATED)
- 184 of 3400 cores would be used in West US.*

Disclaimer: This price estimate does not include storage.

开源 R

```
mydata <- read.csv("http://www.ats.ucla.edu/stat/data/binary.csv")  
  
mylogit <- glm(admit ~ gre + gpa + rank, data = mydata,  
               family = "binomial")
```

R Server

Switch functions

```
mydata <- RxTextData("/data/binary.csv", fileSystem = hdfsFS)
```

```
mylogit <- rxLogit(admit ~ gre + gpa + rank, data = mydata)
```

R Server 通过 Spark 实现并行计算

切换计算上下文

```
rxSetComputeContext( RxSpark(...) )
```

```
mydata <- RxTextData("/data/binary.csv", fileSystem = hdfsFS)
```

```
mylogit <- rxLogit(admit ~ gre + gpa + rank, data = mydata)
```

R Server 通过MapReduce 实现并行计算

切换计算上下文

```
rxSetComputeContext( RxHadoopMR(...) )
```

```
mydata <- RxTextData("/data/binary.csv", fileSystem = hdfsFS)
```

```
mylogit <- rxLogit(admit ~ gre + gpa + rank, data = mydata)
```

新的 Spark 计算要比 Map Reduce 快 3-7 倍

在 R Server 中使用 SparkR

```
#####  
# Join airline data with weather at Origin Airport  
#####  
  
joinedDF <- SparkR::join(  
  airDF,  
  weatherDF,  
  airDF$OriginAirportID == weatherDF$AirportID &  
    airDF$Year == weatherDF$AdjustedYear &  
    airDF$Month == weatherDF$AdjustedMonth &  
    airDF$DayofMonth == weatherDF$AdjustedDay &  
    airDF$CRSDepTime == weatherDF$AdjustedHour,  
  joinType = "left_outer"  
)
```

在 R Server 中训练及评估模型

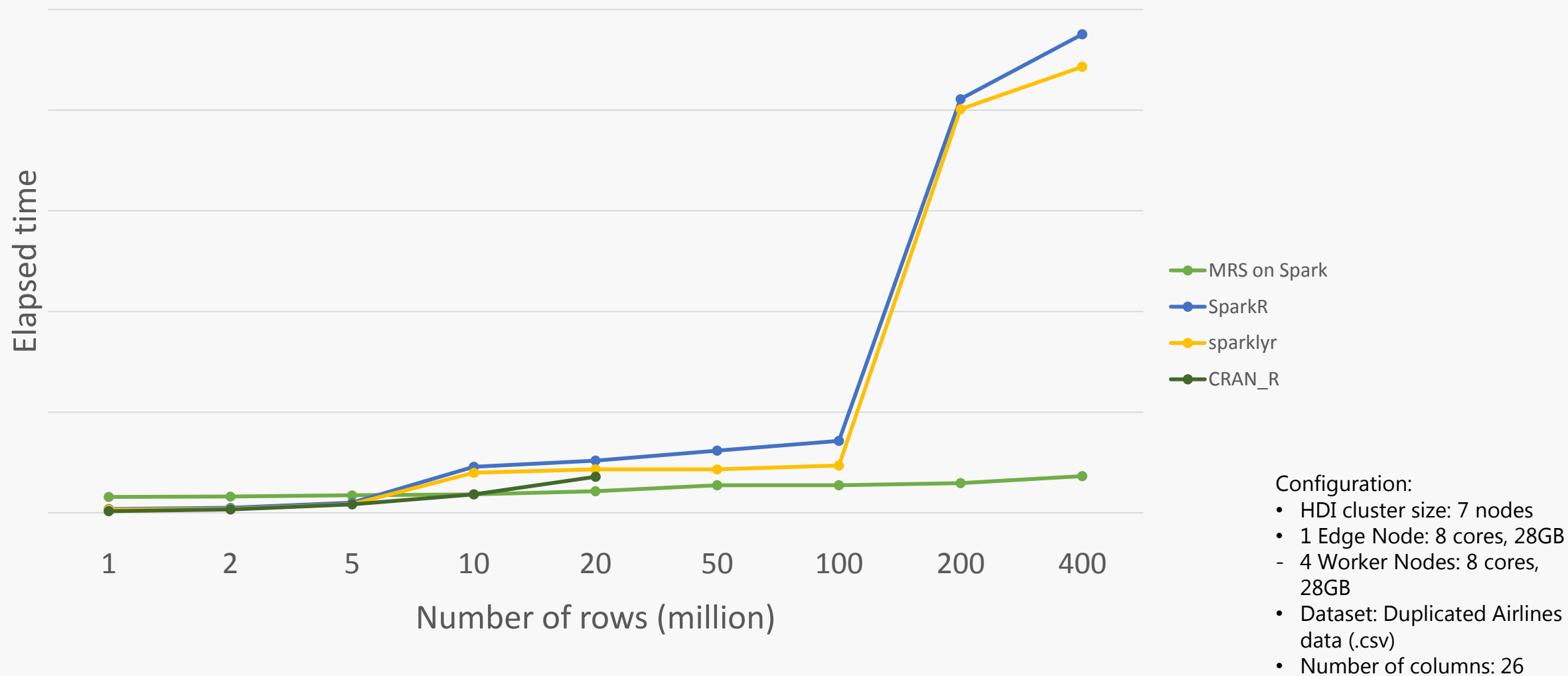
```
#####  
# Train and Test a Decision Tree model  
#####  
  
# Train using the scalable rxDTree function  
  
dTreeModel <- rxDTree(formula, data = trainDS,  
                      maxDepth = 6, pruneCp = "auto")  
  
# Test using the scalable rxPredict function  
  
rxPredict(dTreeModel, data = testDS, outData = treePredict,  
          extraVarsToWrite = c("ArrDel15"), overwrite = TRUE)
```

发布 Web Service

```
#####  
# Publish the scoring function as a web service  
#####  
  
library(AzureML)  
  
workspace <- workspace(config = "azureml-settings.json")  
  
endpoint <- publishWebService(workspace, scoringFn,  
                              name="Delay Prediction Service",  
                              inputSchema = exampleDF)  
  
#####  
# Score new data via the web service  
#####  
  
scores <- consume(endpoint, dataToBeScored)
```

R Server on Spark – 高性能、高伸缩性

Logistic Regression (executing models)



R + 机器学习：助力最佳零售数字化体验

千人千面：打造完美的个性化体验



推荐、促销、导购

满足个性化购买喜好/愿望，随时、随地提供丰富的商品信息/推荐

跨区间销售，实时响应

深入洞察消费者，基于社交/移动的销售协作，实时响应消费者的购买需求/意愿，线上-线下互动

提供个性化消费体验

消费历史追踪，跨渠道整合消费偏好，推动交叉销售/追加销售

解决方案：交叉销售/追加销售

销售商品

- 店面商品
- 库存商品
- 网店商品
- 加盟店商品
- 首发新品
- 促销商品
- 专卖商品
- 折扣商品



顾客画像

- 销售记录
- 客户记录
- 社交信息
- 面部识别

VIP客户

会员客户

新客户

潜在客户

流失客户

销售渠道



现场导购



线上导购



二维码



公众号



邮件/短信

目标

提升消费者体验，通过个性化推荐，
实现交叉营销、追加营销、精准营销

搭建模型 – User-User Collaborative Filtering

Training data

User	Product	Amount
89384	ESTEE LAUDER	40.28
89384	COACH	10.47
89384	CHANEL	4.31
48982	LANCOME	507.40
48982	DIOR	13.54
...	...	...

Scoring data

User	Product	Amount
74937	CHANEL	12.53
74937	DIOR	84.61
74937	SK-II	1.08
20740	DIOR	44.78
20740	ESTEE LAUDER	208.11
...	...	...

Training rating matrix

User	COACH	DIOR	...
89384	3	1	...
48982	1	0	...
...	...	...	...

Usage → Ratings

Rating system = q stars

Bucket training usage data into q quantiles for each product

Then bucket the scoring data using the same cutoffs

Scoring rating matrix

User	COACH	DIOR	...
74937	0	4	...
20740	4	3	...
...	...	...	...

Distance calculator

Calculate distance between every scoring user and every training user

$$D_{ij} = \sum_{k=1}^m |R_{ik} - R_{jk}|$$

Expected rating calculator

Specify k nearest neighbors and do a similarity weighted average for each product, e.g.

$$\hat{R}_{i, \text{sql db}} = \frac{\sum_{j \in knn} e^{-D_{ij}} R_{j, \text{sql db}}}{\sum_{j \in knn} e^{-D_{ij}}}$$

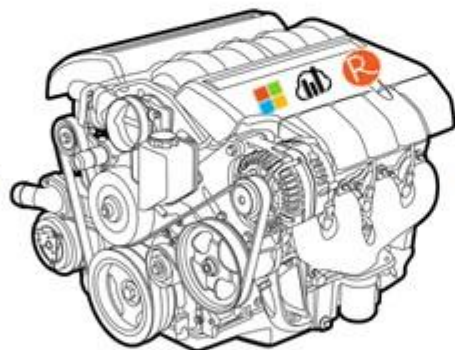
For each customer, pick the N highest scoring products not currently used

实现及运作

业务数据

User	Product	Amount
893848	CHANEL	40.28
893848	DIOR	10.47
893848	SK-II	4.31
489822	DIOR	507.40
489822	ESTEE LAUDER	13.54
...	...	...

推荐引擎



渠道



消费者

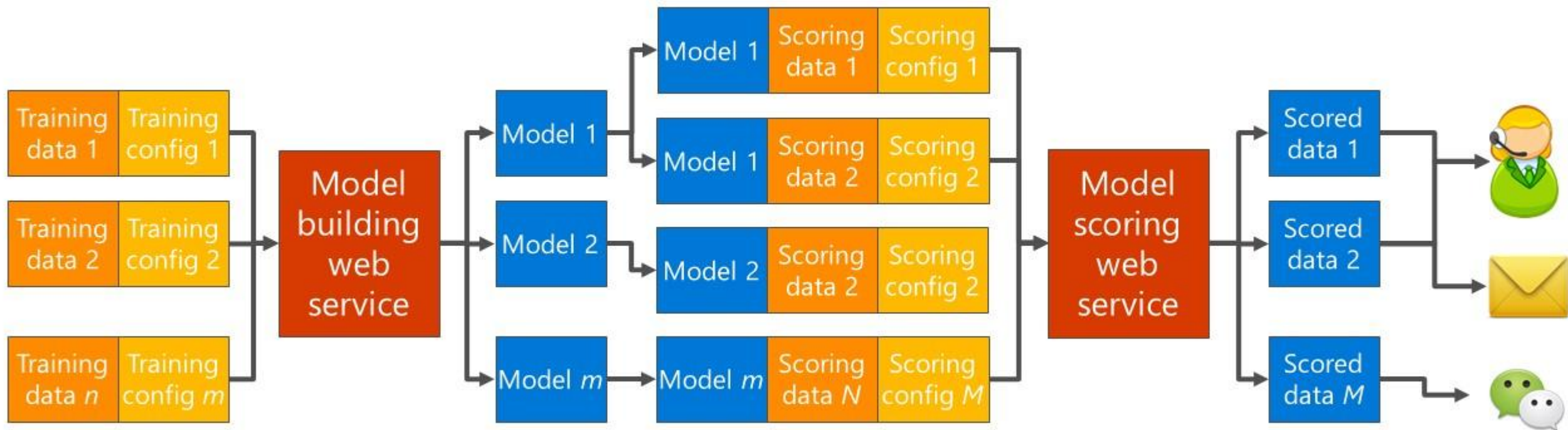


Analytics & Insights

实验配置



多通道的个性化推荐



智胜之道：深入业务洞察， 智能商业模式创新



销量分析预测

订单销量、多层级销量、新品销售
消费者满意度、客户流失情况

市场营销投入

渠道投入、渠道贡献
渠道拆解与重构、渠道协同

物流、库存分析预测

物流流转、区域配送
拆单发货、跨区发货
仓库吞吐、分仓配货补货

案例分析： 市场渠道及回报

超过不同来源与不同格式的四百多个变量才能描述客户所处的真实商业环境。

其中包含所有广告与促销投资及其所产生的影响。

以及不在客户控制范围的外部因素，比如竞品市场活动、宏观经济环境甚至天气。

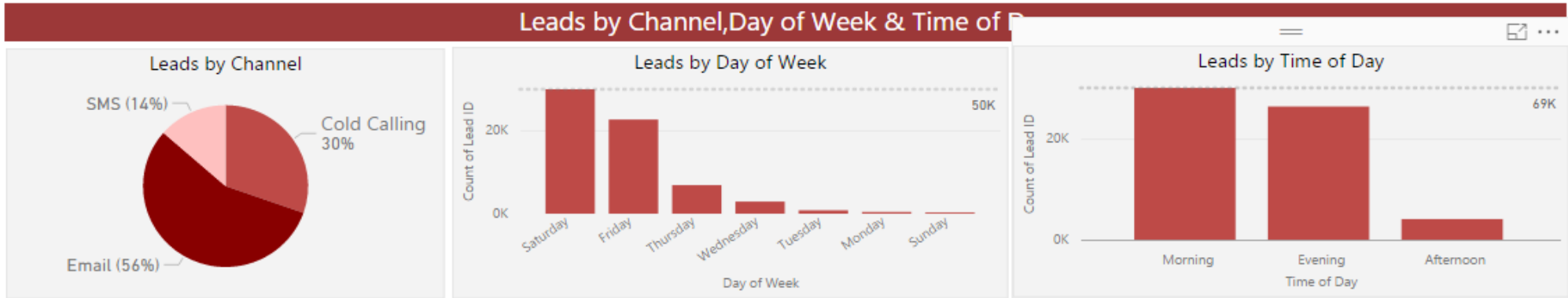
Source: HighCloud MIA Solution on Azure ML, Microsoft Ignite China 2016

		Page Layout	Formulas	Data	Review	View	Add-ins	Team	Analyze	Design	Tell me what you want to do															
ExFactors	Newspaper	131	135	125	111	116	109	124	225	186	206	190	191	23	162	160	160	176	154	144	16					
	Newspaper Cost	131	135	125	111	116	109	124	225	186	206	190	191	23	162	160	160	176	154	144	16					
	Newyear07Dip																									
	Newyear08Peak																									
	Newyear08PeakO																									
	Newyear09Dip																									
	NPMG	185	197	125	111	179	129	166	225	226	206	231	191	23	195	201	160	176	197	165	18					
	NPMG Cost	185	197	125	111	179	129	166	225	226	206	231	191	23	195	201	160	176	197	165	18					
	OtherMedia	52	52	53	52	57	55	55	68	78	62	61	62	100	119	90	90	90	99	99	10					
	OtherMedia Cost	52	52	53	52	57	55	55	68	78	62	61	62	100	119	90	90	90	99	99	10					
	Print	210	213	150	127	203	145	196	241	249	221	264	207	31	234	226	199	200	236	189	22					
	Print Cost	210	213	150	127	203	145	196	241	249	221	264	207	31	234	226	199	200	236	189	22					
	QDM				77	63	27	13	78	64	27	13	82	67	42			68	55	24	11	4				
	QDM Cost				77	63	27	13	78	64	27	13	82	67	42			68	55	24	11	4				
	Search	15	17	23	24	27	20	21	24	22	17	18	15	12	16	16	15	16	18	21	2					
	Search Cost	15	17	23	24	27	20	21	24	22	17	18	15	12	16	16	15	16	18	21	2					
Fujitsu	Building Sales Price Index																									
	Consumer Confidence Index																									
	Consumer Expectation Index																									
	Consumer Price Index																									
	Consumer Satisfaction Index																									
	Exchange Rate																									
	GDP	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###					
	Gross Industrial Output Value	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###					
	Interest Rate (benchmark one-year deposit rates)																									
	Real Interest Rates																									
	Shanghai Stock Exchange Composite Index _High	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2					
	Shanghai Stock Exchange Composite Index _Low	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2					
	Total Retail Sales of Consumer Goods	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###				
HP	Magazine	8			20				7	40				77	2			46								
	Newspaper	6															37	43	48	27	16					
	TV																									
IBM	Magazine																									
	Newspaper																									
	TV																									
Lenovo	Magazine	3	67	58	120	43	78	52	185	9	96	35	48	369	12	3	84	161	10	20	2					
	Newspaper	17	137	224	253	165	185	325	246	260	387	176	197	145	263	200	101	70	40	85	10					
	TV	106	147	158	225	485	###	###	###	828	604	588	588	674	747	494	229	182	108	210	22					
NEC	Magazine								9																	
	Newspaper																									
	TV																									
Samsung	Magazine	18			81		18		77			19	26	90	12	9	9	73	10	33						
	Newspaper	27	43	17	1	4	5	5	5	9	34	3	9	40	32	1	1	1	15	7						
	TV																									
Sony	Magazine		19		15			10	30	9		43	27		9		97		5							
	Newspaper	7											16	14	14		27	14								

案例分析：

市场渠道及回报

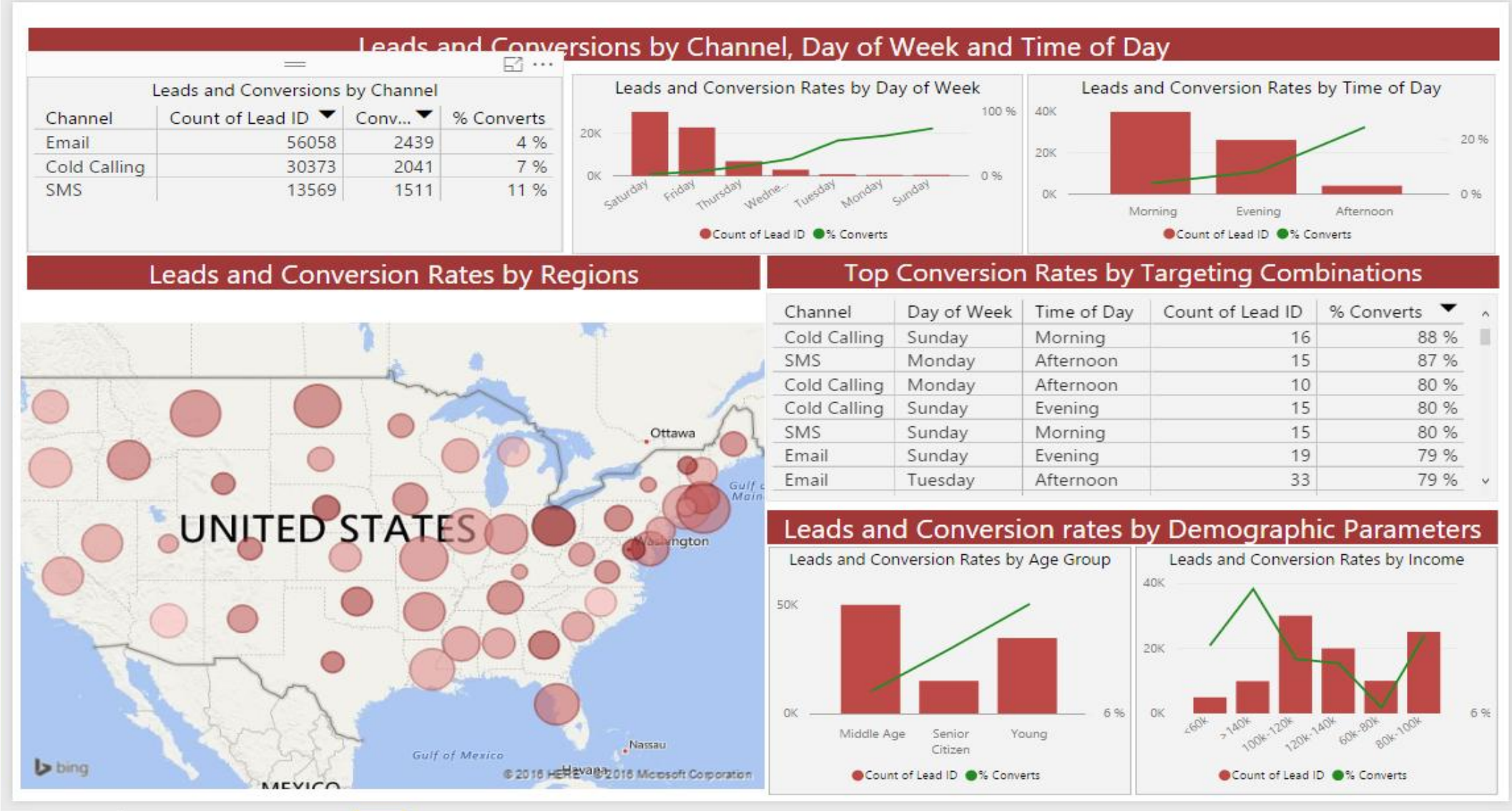
Recommended Channel by Day and Time						
Lead ID	Campaign Name	Product	Recommended Channel	Recommended Day	Recommended Time	Conv Probability
000311ac-7b4a-cb60-6d72-9653de0a9...	Happiness insured	H1	Email	Saturday	Morning	25.60 %
00037b5d-7a7b-1a9a-e8cf-299a47c30...	Know Money	LTC3	Cold Calling	Friday	Evening	16.20 %
000473a7-1094-56ce-2b9b-ba9e3d63...	Happiness insured	H1	Email	Saturday	Morning	18.60 %
00050e64-6734-69f3-0e2d-ae6e101cc...	New dawn	D3	Email	Saturday	Afternoon	17.80 %
0005277e-a862-66e2-91cb-da02c8861...	New dawn	D3	SMS	Thursday	Morning	17.20 %
0005324a-48dd-7378-f565-98183b156...	Protect your future	D2	Cold Calling	Saturday	Morning	16.20 %



Leads by Demographic Parameters		Top Recommended Combinations																																					
Leads by Channel and Age group	Leads by Channel and Income Bucket	<table><tr><th>Recommended Channel</th><th>Recommended Day</th><th>Recommended Time</th><th>Count of Lead ...</th></tr><tr><td>Email</td><td>Saturday</td><td>Morning</td><td>27685</td></tr><tr><td>Cold Calling</td><td>Saturday</td><td>Morning</td><td>13994</td></tr><tr><td>Email</td><td>Saturday</td><td>Evening</td><td>10207</td></tr><tr><td>Email</td><td>Friday</td><td>Morning</td><td>8096</td></tr><tr><td>SMS</td><td>Saturday</td><td>Morning</td><td>5742</td></tr><tr><td>Cold Calling</td><td>Saturday</td><td>Evening</td><td>5572</td></tr><tr><td>Cold Calling</td><td>Friday</td><td>Morning</td><td>4181</td></tr><tr><td>Email</td><td>Friday</td><td>Evening</td><td>3187</td></tr></table>	Recommended Channel	Recommended Day	Recommended Time	Count of Lead ...	Email	Saturday	Morning	27685	Cold Calling	Saturday	Morning	13994	Email	Saturday	Evening	10207	Email	Friday	Morning	8096	SMS	Saturday	Morning	5742	Cold Calling	Saturday	Evening	5572	Cold Calling	Friday	Morning	4181	Email	Friday	Evening	3187	
Recommended Channel	Recommended Day	Recommended Time	Count of Lead ...																																				
Email	Saturday	Morning	27685																																				
Cold Calling	Saturday	Morning	13994																																				
Email	Saturday	Evening	10207																																				
Email	Friday	Morning	8096																																				
SMS	Saturday	Morning	5742																																				
Cold Calling	Saturday	Evening	5572																																				
Cold Calling	Friday	Morning	4181																																				
Email	Friday	Evening	3187																																				

案例分析：

市场渠道及回报

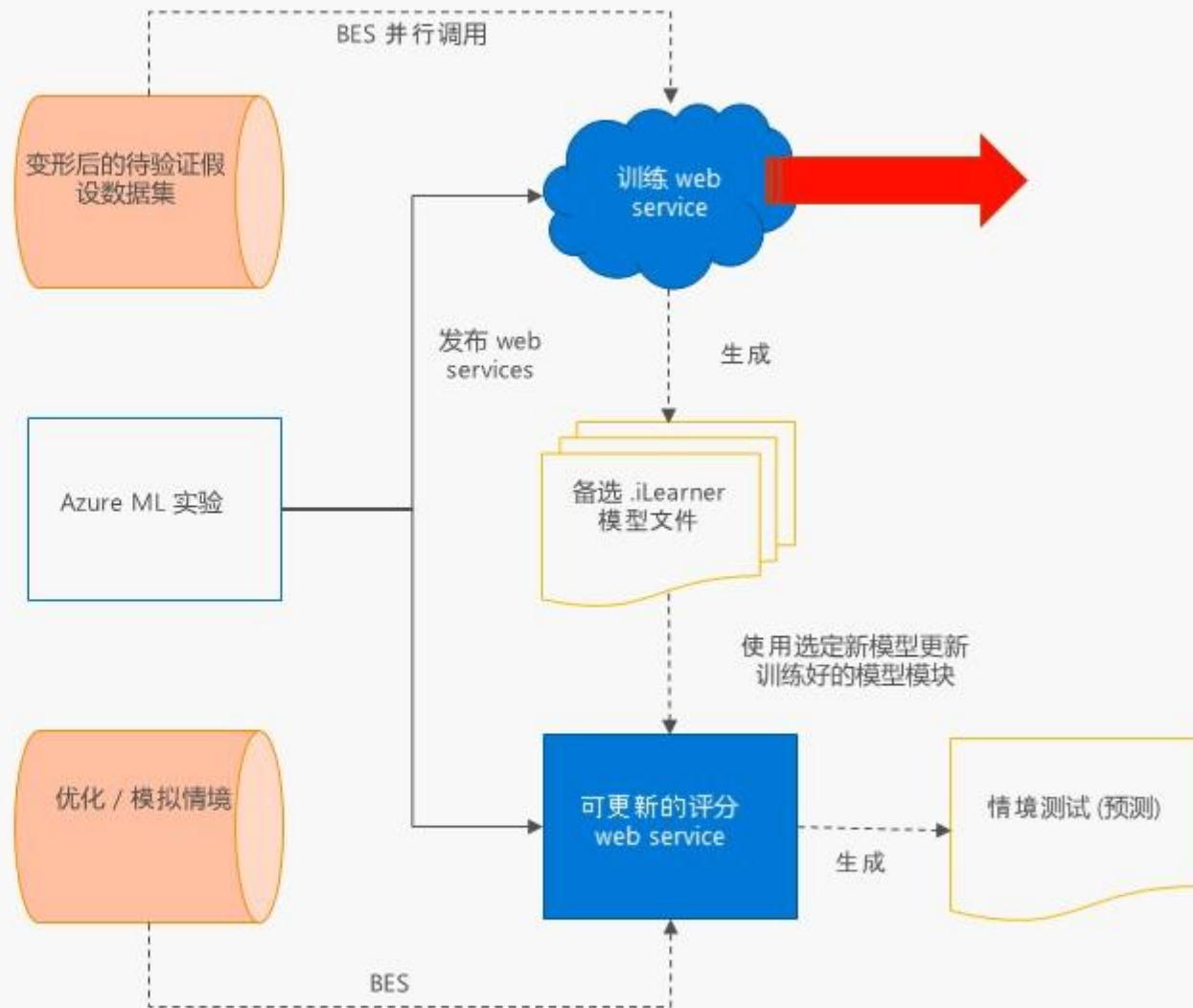


案例分析： 市场渠道及回报

每endpoint最高200个并发调用
每秒4000次回归运算(50ms任务)
可通过增加endpoints扩展计算能力

VS.

传统统计软件每秒10-30次回归运算



总结

最佳体验最关键

消费者：全面的线上线下个性化体验

零售商：深入业务洞察，商业模式创新

Hadoop + Spark + R

微软云高级分析服务及 R 平台

Hadoop, R and Spark are better together

Performance, Scalability are critical

学 R 以致用，广纳 Open 之源

使用 R 打造端到端的大数据应用

使用 R，为 R 开源社区奉献

开发推动零售行业数字化转型的ML算法

谢谢！