

Contents

1	Introduction to R for Cloud Computing	1
1.1	What Is R Really?	1
1.2	What Is Cloud Computing?	2
1.2.1	What Is SaaS, PaaS, and IaaS?	3
1.3	Why Should Cloud Users Learn More About R?	4
1.4	Big Data and R	5
1.5	Why Should R Users Learn More About the Cloud?	6
1.6	How Can You Use R on the Cloud Infrastructure?	6
1.7	How Long Does It Take to Learn and Apply R on a Cloud?	7
1.8	Commercial and Enterprise Versions of R	7
1.9	Other Versions of R	9
1.10	Cloud Specific Packages of R	10
1.11	Examples of R and Cloud Computing	10
1.12	How Acceptable Is R and the Cloud to Enterprises?	12
1.12.1	SAP Hana with R Interview 1 (7 June 2012)	12
1.12.2	SAP Hana with R Interview 2–14 June 2012	14
1.12.3	TIBCO TERR	14
1.13	What Is the Extent of Data that Should Be on Cloud Versus Local Machines?	16
1.14	Alternatives to R	17
1.15	Interview of Prof Jan De Leeuw, Founder of JSS	18
2	An Approach for Data Scientists	21
2.1	Where to Get Data?	22
2.2	Where to Process Data?	23
2.2.1	Cloud Processing	24
2.3	Where to Store Data?	24
2.3.1	Cloud Storage	25
2.3.2	Databases on the Cloud	26
2.4	Basic Statistics for Data Scientists	27
2.5	Basic Techniques for Data Scientists	28

2.6	How to Store Output?	28
2.7	How to Share Results and Promote Yourself?.....	29
2.8	How to Move or Query Data?.....	29
2.8.1	Data Transportation on the Cloud	29
2.9	How to Analyse Data?.....	30
2.9.1	Data Quality—Tidy Data	30
2.9.2	Data Quality—Treatment	30
2.9.3	Data Quality—Transformation	31
2.9.4	Split Apply Combine	31
2.10	How to Manipulate Data?	31
2.10.1	Data Visualization.....	32
2.11	Project Methodologies.....	32
2.12	Algorithms for Data Science	33
2.13	Interview of John Myles White, co-author of Machine Learning for Hackers	36
3	Navigating the Choices in R and Cloud Computing	37
3.1	Which Version of R to Use?.....	37
3.1.1	Renjin.....	37
3.1.2	pqR	37
3.2	Which Interface of R to Use	38
3.3	Using R from the Browser	40
3.3.1	Statace	40
3.3.2	R Fiddle	42
3.3.3	RShiny and RStudio	45
3.4	The Cloud Computing Services Landscape	48
3.4.1	Choosing Infrastructure Providers	49
3.4.1.1	Amazon Cloud	49
3.4.1.2	Other Components of Amazon Cloud	50
3.4.1.3	Google Cloud Services	51
3.4.1.4	Windows Azure	53
3.5	Interview Ian Fellows Deducer	54
3.6	Notable R Projects	57
3.6.1	Installr Package	57
3.6.2	Rserve	58
3.6.3	RApache	58
3.6.4	Rook	58
3.6.5	RJ and Rservi.....	58
3.6.6	R and Java	59
3.7	Creating Your Desktop on the Cloud	59

4	Setting Up R on the Cloud	61
4.1	Connecting to the Cloud Instance Easily	62
4.2	Setting Up R in Amazon's Cloud	64
4.2.1	Local Computer (Windows) Cloud (Linux)	64
4.3	Using Revolution R on the Amazon Cloud	75
4.4	Using the BioConductor Cloud AMI	76
4.5	R in Windows Azure Cloud	80
4.6	R in the Google Cloud	93
4.6.1	Google Big Query	105
4.6.2	Google Fusion Tables API	105
4.6.3	Google Cloud SQL	107
4.6.4	Google Prediction API	107
4.7	R in IBM SmartCloud	107
4.7.1	IBM SmartCloud Enterprise	107
4.7.2	IBM Softlayer	124
4.7.3	Using R with IBM Bluemix	125
5	Using R	127
5.1	A R Tutorial for Beginners	127
5.2	Doing Data Mining with R	131
5.3	Web Scraping Using R	142
5.4	Social Media and Web Analytics Using R	144
5.4.1	Facebook Network Analytics Using R	144
5.4.2	Twitter Analysis with R	149
5.4.3	Google Analytics API with R	150
5.4.4	R for Web Analytics	157
5.5	Doing Data Visualization Using R	157
5.5.1	Using Deducer for Faster and Better Data Visualization in R	158
5.5.2	R for Geographical Based Analysis	170
5.6	Creating a Basic Forecasting Model with R	174
5.7	Easy Updating R	176
5.8	Other R on the Web Projects	177
5.8.1	Concerto	177
5.8.2	Rapporter	179
5.8.3	R Service Bus	186
5.9	The Difference Between Open Source and Closed Source Enterprise Software	188

6	Using R with Data and Bigger Data	193
6.1	Big Data Interview	193
6.2	The Hadoop Paradigm	197
6.2.1	What Is Hadoop All About	197
6.2.2	The Ecosystem of Hadoop	197
6.2.3	Current Developments in Hadoop	200
6.3	Commercial Distributions of Hadoop	201
6.3.1	Hadoop on the Cloud as a Service	202
6.3.2	Learning Hadoop	202
6.4	Learning Hadoop	202
6.5	RHadoop and Other R Packages	203
6.6	Interview with Antonio Piccolboni, Consultant on RHadoop	204
6.7	Big Data R Packages	206
6.8	Databases and CAP Theorem	208
6.8.1	CAP Theorem Visually	209
6.9	ACID and BASE for Databases	209
6.10	Using R with MySQL	210
6.11	Using R with NoSQL	211
6.11.1	MongoDB	211
6.11.2	jsonlite (prettify)	213
6.11.3	CouchDB	214
6.11.4	MonetDB	214
6.12	Big Data Visualization	215
7	R with Cloud APIs	217
7.1	Everything Has an API	217
7.2	REST	218
7.3	rOpenSci	219
7.4	What Is Curl	221
7.5	Navigating OAuth2	221
7.6	Machine Learning as a Service	222
7.6.1	Google Prediction API and R	222
7.6.2	BigML with R	223
7.6.3	Azure Machine Learning with R	223
7.7	Examples of R with APIs	223
7.7.1	Quandl with R	223
7.7.2	Data Visualization APIs with R	224
7.7.2.1	Plotly	224
7.7.2.2	googleVis	227
7.7.2.3	tabplotd3	228
7.7.2.4	Yhat and R	229
7.7.3	OpenCPU and R	229
7.7.3.1	Interview Jeroen Ooms OpenCPU	230
7.8	RForcecom by Takekatsu Hiramura	232

8 Securing Your R cloud	237
8.1 Ensuring R Code Does not Contain Your Login Keys	237
8.2 Setting Up Access Control (User Management Rights)	238
8.2.1 Amazon User Management.....	238
8.3 Setting Up Security Control Groups for IP Address Level Access (Security Groups)	240
8.3.1 A Note on Passwords and Passphrases	241
8.3.2 A Note on Social Media's Impact on Cyber Security	242
8.4 Monitoring Usage for Improper Access	243
8.5 Basics of Encryption for Data Transfer (PGP- Public Key, Private Key)	243
8.5.1 Encryption Software	243
9 Training Literature for Cloud Computing and R	247
9.1 Challenges and Tips in Transitioning to R	247
9.2 Learning R for Free Online	252
9.3 Learn More Linux	253
9.4 Learn Git	254
9.5 Reference Cards for Data Scientists	254
9.6 Using Public Datasets from Amazon	255
9.7 Interview Vivian Zhang Co-founder SupStat.....	256
9.8 Interview EODA	258
9.9 Rpubs	260
Appendix	261
A.1 Creating a Graphical Cloud Desktop on a Linux OS	261
References	263

<http://www.springer.com/978-1-4939-1701-3>

R for Cloud Computing

An Approach for Data Scientists

Ohri, A.

2014, XVII, 267 p. 255 illus., 160 illus. in color.,

Softcover

ISBN: 978-1-4939-1701-3