

Pro Salesforce Analytics Cloud

A Guide to Wave Platform,
Builder, and Explorer

—

William Smith
Helen Sun

Apress®

Pro Salesforce Analytics Cloud

A Guide to Wave Platform, Builder,
and Explorer



William Smith

Helen Sun

Apress®

Pro Salesforce Analytics Cloud: A Guide to Wave Platform, Builder, and Explorer

Copyright © 2016 by William Smith and Helen Sun

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.

ISBN-13 (pbk): 978-1-4842-1204-2

ISBN-13 (electronic): 978-1-4842-1203-5

Trademarked names, logos, and images may appear in this book. Rather than use a trademark symbol with every occurrence of a trademarked name, logo, or image we use the names, logos, and images only in an editorial fashion and to the benefit of the trademark owner, with no intention of infringement of the trademark.

The use in this publication of trade names, trademarks, service marks, and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Managing Director: Welmoed Spahr

Acquisitions Editor: Susan McDermott

Technical Reviewers: Mike Martin, Carl Brundage

Editorial Board: Steve Anglin, Pramila Balen, Louise Corrigan, James DeWolf, Jonathan Gennick,

Robert Hutchinson, Celestin Suresh John, Michelle Lowman, James Markham, Susan McDermott,

Matthew Moodie, Jeffrey Pepper, Douglas Pundick, Ben Renow-Clarke, Gwenan Spearing

Coordinating Editor: Rita Fernando

Copy Editor: Carole Berglie

Compositor: SPi Global

Indexer: SPi Global

Cover Designer: Friedhelm Steinen-Broo

Distributed to the book trade worldwide by Springer Science+Business Media New York, 233 Spring Street, 6th Floor, New York, NY 10013. Phone 1-800-SPRINGER, fax (201) 348-4505, e-mail orders-ny@springer-sbm.com, or visit www.springer.com. Apress Media, LLC is a California LLC and the sole member (owner) is Springer Science + Business Media Finance Inc (SSBM Finance Inc). SSBM Finance Inc is a Delaware corporation.

For information on translations, please e-mail rights@apress.com, or visit www.apress.com.

Apress and friends of ED books may be purchased in bulk for academic, corporate, or promotional use. eBook versions and licenses are also available for most titles. For more information, reference our Special Bulk Sales-eBook Licensing web page at www.apress.com/bulk-sales.

Any source code or other supplementary materials referenced by the author in this text is available to readers at www.apress.com. For detailed information about how to locate your book's source code, go to www.apress.com/source-code/.

For Nathaniel G. Smith

Contents at a Glance

About the Authors	xiii
About the Technical Reviewers	xv
Acknowledgments	xvii
Introduction	xix
■ Part I: A Complete Guide to the Salesforce Analytics Cloud	1
■ Chapter 1: An Overview of the Salesforce Analytics Cloud	3
■ Chapter 2: The Salesforce Analytics Cloud Explorer	15
■ Chapter 3: Analytics Cloud Builder.....	39
■ Chapter 4: Salesforce Analytics Platform.....	77
■ Part II: Business Analytics Solutions Using the Salesforce Analytics Cloud	87
■ Chapter 5: Critical Decision Making and the Salesforce Analytics Cloud	89
■ Chapter 6: Mobile Enterprise Data Discovery	111
■ Chapter 7: Advanced Data Acquisition and Processing	141
Index	157

Contents

About the Authors	xiii
About the Technical Reviewers	xv
Acknowledgments	xvii
Introduction	xix
■ Part I: A Complete Guide to the Salesforce Analytics Cloud	1
■ Chapter 1: An Overview of the Salesforce Analytics Cloud	3
Evolution of Business Intelligence and Reporting Systems	3
The Challenge of Data Variety	5
Data Uncertainty	5
The Scarcity of Personnel.....	6
Data Discovery	7
Self-Service Application Development and Analytic Interface.....	8
A Data Server.....	8
Data-Loading Interfaces and APIs	8
An Alternative to the Data Lake	8
Introduction to the Salesforce Analytics Cloud.....	9
The Mobile Platform	9
The Desktop Platform	10
Advantages of the Salesforce Analytics Cloud	12
The Cloud Architecture	12
Mobile First.....	12
Integration with the Salesforce Ecosystem	12
Conclusion.....	13

- **Chapter 2: The Salesforce Analytics Cloud Explorer 15**
- What Is Cloud Explorer? 15**
 - The Two Versions of the Salesforce Analytics Cloud..... 15
- Getting Started with Cloud Explorer 16**
 - Launching Cloud Explorer 16
- Datasets 19**
 - Row-Level Security for Datasets 19
- Lenses 19**
 - Measures 19
 - Dimensions 20
 - Dates 22
 - Bringing It All Together 23
 - Learn by Doing..... 26
 - Lens Modification History, Undo, and Replay..... 27
 - Exploring the Data 27
 - Mobile Lens Creation 28
- Dashboards 28**
 - Widgets..... 28
 - Data Display Widgets..... 29
 - Selector Widgets..... 30
 - Display Widgets 31
 - The Dashboard Development Cycle 32
 - Link Back to Lens 34
 - Dashboard Wrap-Up 36
- Apps 36**
 - My Private App..... 36
 - The Share Button..... 36
 - Playground Data 37

What Wave Lacks	37
Text Processing and Analysis	38
Map and Geospatial Data.....	38
Advanced Math and Statistical Functions	38
Conclusion.....	38
■ Chapter 3: Analytics Cloud Builder.....	39
Building a Simple Dashboard	39
Loading Dataset.....	39
Creating a Dashboard.....	44
Dashboard Design Aesthetics.....	54
Steps to Design Aesthetic Dashboards.....	55
Mobile Dashboard Designer	56
Upload Dataset Through Mobile Device.....	56
Create Dashboard Using Mobile Designer	64
Advanced Dashboards.....	70
Anatomy of a dashboard.....	70
Advanced Dashboard with SAQL	71
Summary.....	75
■ Chapter 4: Salesforce Analytics Platform.....	77
Methods of Data Integration.....	77
External Data API.....	79
Step 1. Create .CSV File.....	80
Step 2. Create JSON Metadata	81
Step 3. Upload Files.....	81
Step 4. System Job Is Created and Run	81
Step 5. Dataset Is Created.....	81
Restrictions on Use.....	81
Supporting Documents.....	82

Extract Transform Load Tools	82
Step 1. Connect to Source	83
Step 2. Extract Data.....	84
Step 3. Sort, Filter, and Transform	84
Step 4. Determine the Dataset	84
Step 5. Write the Dataset.....	84
Wave REST API	85
Conclusion.....	85
■ Part II: Business Analytics Solutions Using the Salesforce Analytics Cloud	87
■ Chapter 5: Critical Decision Making and the Salesforce Analytics Cloud	89
Critical Decision Making and Cognitive Bias	89
The Use Case Manufacturer	90
Early Glass Making	91
Early Instrumentation and Increased Statistical Process Control.....	92
Automatic Control via Computer Technology	92
Data Networks Offer Increased Integration	92
Company Structure and Facilities	94
Float Glass Plants	94
Glass Fabrication Plants	94
Automotive Encapsulation Facilities.....	94
The Organization of the Company's Plants	95
Key Performance Indicators	96
Facility-Level Reporting	96
Diagnosis of Machine Problem	100
Eliminating Cognitive Bias with Data.....	103
Corporate-Level Reporting	104
Assessing Plant Manager Performance.....	106
Conclusion.....	110

■ Chapter 6: Mobile Enterprise Data Discovery	111
Overview of Our Use Case	111
The Traditional Approach to Reporting.....	112
An Alternative to the Traditional Approach.....	112
The Use Case Architecture	114
Step 1 – Output to .CSV File	115
Step 2 – User Access of .CSV File on Dropbox	116
Step 3 – Modifying the Metadata and Creating the Dataset.....	119
Step 4 – Creating Lenses and Dashboards.....	122
Data Used in Our Use Case.....	122
Sales Amount Data	123
Date Data	123
Product Data	123
Reseller Data	123
Dashboard Development Cycle	124
Digging Deeper	128
Adding Other Dimensions: Evaluating Individual Contributors	136
Analyzing Other Dimensions: Resellers	138
Security and Shared Public File Services.....	139
Conclusion.....	140
■ Chapter 7: Advanced Data Acquisition and Processing	141
Data Acquisition—Tools and Technologies	141
Web Scrapers	141
Web Crawlers	142
Text Analytics Processors	142
Applications of These Tools	142

Data Processing—Terms and Concepts.....	145
Data Wrangling	145
Data Standards.....	145
Data Quality	145
Data Cleansing.....	146
Python: A Programming Language for Data Processing.....	146
Two Examples Using Python.....	148
Data Cleansing Example 1	148
Data Cleansing Example 2.....	150
Python Resources.....	154
Conclusion.....	155
Index.....	157

About the Authors

William Smith is Chief Architect at Vendita Technology Group, and where he is in charge of the Advanced Product Group. William is patented inventor of engineered systems for database and analytic systems, based on a converged infrastructure featuring both RISC and X86 processors. He is co-author of *Master Competitive Analytics with Oracle Endeca Information Discovery* (McGraw-Hill/Oracle Press) and has more than two decades experience as an enterprise architect with strong background in decision support systems and analytics. William's early career experience includes working with data from manufacturing control systems and instrumentation in industries that include plastics, petro-chemicals, glass, food, and extrusion processes. William has enterprise systems experience in higher education, finance, consumer research, and healthcare. His strong database administration and programming skills have enabled him to address analytical systems with a comprehensive and complete set of technical knowledge, resulting in top down designed systems and architectures. In the area of analytics, William began with working in the area of predictive failures and statistical process control. Later projects included utilization of advanced analytics technologies to provide insight in consumer research, institutional research, clinical informatics, and finance. William is currently working with customers making use of Big Data and is developing innovative solutions utilizing Salesforce Analytics Cloud.

Helen Sun, PhD, is the Vice President of Enterprise Architecture and Cloud Computing at Motorola Solutions with over 16 years of business and technology leadership experience in various industries including financial services, retail, CPG, logistics, healthcare, telecommunications, and utilities and with deep experience and proven expertise in analytics solutions and enterprise architecture. She is the co-author of *Oracle Big Data Handbook and Master Competitive Analytics with Oracle Endeca Information Discovery* (McGraw-Hill/Oracle Press). Helen is considered one of the Big Data luminaries in the industry.

About the Technical Reviewers



Mike Martin is a Solution Architect at Appirio where he is focused on helping Salesforce customers get the most out of their investment in the platform by scoping implementations and designing custom solutions for customers. Mike has completed the Salesforce Wave Analytics Cloud Brown Belt Accreditation and is a Certified Salesforce Administrator, Advanced Administrator, Sales Cloud Consultant, Service Cloud Consultant, Force.com Developer, and Platform App Builder. Mike has also been recognized as a Salesforce Community MVP and helps to lead the Indianapolis Salesforce User Group. He has been active in the Salesforce community since 2008.

Mike is a graduate of Rose-Hulman Institute of Technology with a BS in Computer Engineering and a Masters of Engineering Management. Mike currently lives in Indianapolis, IN with his wife, Jennifer, and two sons, Alexander and Jameson. When not working with Salesforce, you can probably find Mike enjoying a craft beer or helping with media and tech production at church. You can find Mike on Twitter: @mikemartin__c.



Carl Brundage is passionate about data, analytics and software's ability to solve business challenges. Throughout his career, he has been a creator, whether it be technologies, departments, practices, products or customer success. Carl excels at the intersection of business and technology, where there is a need to create something new. He has delivered multiple presentations and blog posts on the Salesforce Analytics Cloud and earned Brown Belt Accreditation as an early adopter.

Recently, as a chief product evangelist, Carl does everything required to launch new products to the Salesforce App Exchange. From creating product roadmaps and developing marketing strategies to writing implementation guides and delivering demos, there is no job too big or small to tackle. In addition, he is a Salesforce certified Sales Cloud Consultant, Service Cloud Consultant, Pardot Consultant, Force.com

Developer, Platform App Builder, Administrator, Advanced Administrator and Marketing Cloud Email Specialist. Connect with Carl on LinkedIn (<https://www.linkedin.com/in/cbrundage>) or follow him on Twitter (@carlbrundage).

Acknowledgments

We wish to gratefully acknowledge the assistance of our friend and colleague Paul Carlstoem in introducing us to the professionals at Apress. We would like to extend our gratitude to Susan McDermott and Rita Fernando at Apress for working so diligently on this title. Working with the team at Apress has been a pleasure and without their assistance, this book would not have been possible. We wish to extend our thanks to our technical editors, Mike Martin and Carl Brundage. We would like to acknowledge the assistance of Ciolek Ltd, Attorneys at Law in legal matters associated with our efforts as authors. Last, but not least, we wish to acknowledge the support of family and friends in their support our efforts as authors, especially our longtime friend and mentor Dr. Mesbeh Ahmed.

Introduction

Data discovery products are a category of analytics and business intelligence software that dominates new investment and has been gaining popularity, as organizations strive to enable decision-making with data. The reason for this popularity is not hard to understand. With data discovery software, everyday users can create applications that empower them as never before. Users are no longer dependent on their IT departments for reporting and data analysis needs. Data discovery users become “citizen data scientists,” performing basic data analysis and finding correlations. This was previously possible only with high-end statistical analysis packages.

The Salesforce Analytics Cloud is a data discovery product introduced at the Salesforce conference Dreamforce in 2014, and has features that make it a compelling offering for organizations considering an investment in this exciting area of technology. It is 100% cloud-based, and adheres to Salesforce’s strategy of “mobile first,” with a fully functional mobile platform for creating and using data discovery applications. The pricing model for Salesforce Analytics Cloud allows organizations to adopt it with no capital investment. The Analytics Cloud is flat-fee subscription based, and is not subject to data transfer costs. Although the Salesforce Analytics Cloud is oriented towards everyday users, it also has advanced features that enable sophisticated applications to be developed by programmers. Support is available on a 24/7 basis and is included in some of the subscription packages.

Salesforce Analytics Cloud has built-in integration capabilities for building Salesforce Analytics Cloud applications with Salesforce data. It also has the capability of creating applications with data originating outside of Salesforce. This capability—creating applications with outside data—is the focus of this title.

Part I of this title provides guidance on using the Salesforce Analytics Cloud, introducing basic concepts, followed by more sophisticated usage.

Part II focuses on two use cases. The first use case features data from a corporation engaged in manufacturing. It has data from a wide variety of venues, from the factory floor to the executive suite. The second use case chapter illustrates how Salesforce Analytics Cloud is used with sales performance data, and is entirely developed on the Salesforce Analytics Cloud mobile interface. The final chapter covers advanced data acquisition and data processing, and covers how web scrapers and crawlers can be used to harvest data for use in Salesforce Analytics Cloud. This is followed by an overview of data processing, with a discussion on the Python programming language and how it can be used for data processing.

PART I



A Complete Guide to the Salesforce Analytics Cloud

CHAPTER 1



An Overview of the Salesforce Analytics Cloud

Know where to find information and how to use it; that is the secret of success.

—Albert Einstein

Einstein's sentiment, expressed in this concise quotation, captures our innate need to acquire and comprehend information. This is because information enables us to make decisions that are not impaired by bias or poor judgment. One could argue that this thought gave rise to computers, which were first developed to perform mathematical calculations faster. As computers have evolved, innovative strategies and technologies have also evolved to allow information in those computers to be more readily used in decision making. These strategies range from statistical methods to technologies that facilitate the storage of data for rapid retrieval, as well as visualization technologies.

In this book, we cover the Salesforce Analytics Cloud, one of the latest technologies to enable the productive use of information. The Salesforce Analytics Cloud belongs to a class of products known as “data discovery” products; this is the term that is usually used by research firms dealing with this technology. The Salesforce Analytics Cloud features a high-performance storage infrastructure that facilitates rapid data retrieval and advanced visualizations allowing information to be quickly understood, all delivered from the cloud infrastructure that underpins the success of Salesforce products. Before we delve into the Salesforce Analytics Cloud, though, let's set the context for this product and survey the data discovery products with a discussion of computers and technology in reporting and analytics.

Evolution of Business Intelligence and Reporting Systems

As computers evolved into systems to manage information, they were increasingly utilized in business and commerce. These early computers stored data in simple flat files. With increasing volumes of data, a more efficient and organized means of storing the data became necessary, which led to the creation of relational databases. Relational databases were capable of storing transactional data efficiently, but not in a manner that is easily queried or mined. The need to analyze information in transactional databases easily and quickly for reporting and analysis summoned into existence data warehouses and business intelligence (BI) tools. For over 20 years, these warehouses and tools have delivered reliable answers to typical questions, and they will continue to be used for the foreseeable future.

During this time, computer technology progressed at an astonishing rate, owing to increases in the power of microprocessors, the central “brain” of all computers. The 1980s marked the beginning of personal computer usage in business, with the introduction of the IBM personal computer in 1981, utilizing an Intel 8088 microprocessor with a total of 29,000 transistors (transistors are the electronic switches used in microprocessors). By the end of the 1980s, Intel Corporation had introduced the 80486,