
# Getting Started with Amazon Redshift

Enter the exciting world of Amazon Redshift for big data, cloud computing, and scalable data warehousing

Stefan Bauer

[PACKT] PUBLISHING enterprise professional expertise distilled

# Table of Contents

Index

# Getting Started with Amazon Redshift

# Getting Started with Amazon Redshift



First published: June 2013

Production Reference: 1030613

Published by Packt Publishing Ltd.

Livery Place

35 Livery Street

Birmingham B3 2PB, UK.

ISBN 978-1-78217-808-8

www.packtpub.com

Cover Image by Suresh Mogre (<suresh.mogre.99@gmail.com>)

# Credits

**Author**

Stefan Bauer

**Reviewers**

Koichi Fujikawa

Matthew Luu

Masashi Miyazaki

**Acquisition Editors**

Antony Lowe

Erol Staveley

**Commissioning Editor**

Sruthi Kutty

**Technical Editors**

Dennis John

Dominic Pereira

**Copy Editors**

Insiya Morbiwala

Alfida Paiva

**Project Coordinator**

Sneha Modi

**Proofreader**

Maria Gould

**Indexer**

Tejal Soni

**Graphics**

Abhinash Sahu

**Production Coordinator**

Pooja Chiplunkar

**Cover Work**

Pooja Chiplunkar

# About the Author

**Stefan Bauer** has worked in business intelligence and data warehousing since the late 1990s on a variety of platforms in a variety of industries. Stefan has worked with most major databases, including Oracle, Informix, SQL Server, and Amazon Redshift as well as other data storage models, such as Hadoop. Stefan provides insight into hardware architecture, database modeling, as well as developing in a variety of ETL and BI tools, including Integration Services, Informatica, Analysis Services, Reporting Services, Pentaho, and others. In addition to traditional development, Stefan enjoys teaching topics on architecture, database administration, and performance tuning. Redshift is a natural extension fit for Stefan's broad understanding of database technologies and how they relate to building enterprise-class data warehouses.

I would like to thank everyone who had a hand in pushing me along in the writing of this book, but most of all, my wife Jodi for the incredible support in making this project possible.

# About the Reviewers

**Koichi Fujikawa** is a co-founder of Hapyrus a company providing web services that help users to make their big data more valuable on the cloud, and is currently focusing on Amazon Redshift. This company is also an official partner of Amazon Redshift and presents technical solutions to the world.

He has over 12 years of experience as a software engineer and an entrepreneur in the U.S. and Japan.

> To review this book, I thank our colleagues in Hapyrus Inc., Lawrence Gryseels and Britt Sanders. Without cooperation from our family, we could not have finished reviewing this book.

**Matthew Luu** is a recent graduate of the University of California, Santa Cruz. He started working at Hapyrus and has quickly learned all about Amazon Redshift.

> I would like to thank my family and friends who continue to support me in all that I do. I would also like to thank the team at Hapyrus for the essential skills they have taught me.

**Masashi Miyazaki** is a software engineer of Hapyrus Inc. He has been focusing on Amazon Redshift since the end of 2012, and has been developing a web application and Fluent plugins for Hapyrus's FlyData service.

His background is in the Java-based messaging middleware for mission critical systems, iOS application for iPhone and iPad, and Ruby scripting.

His URL address is http://mmasashi.jp/.

# www.PacktPub.com

## Support files, eBooks, discount offers and more

You might want to visit www.PacktPub.com for support files and downloads related to your book.

Did you know that Packt offers eBook versions of every book published, with PDF and ePub files available? You can upgrade to the eBook version at www.PacktPub.com and as a print book customer, you are entitled to a discount on the eBook copy. Get in touch with us at <service@packtpub.com> for more details.

At www.PacktPub.com, you can also read a collection of free technical articles, sign up for a range of free newsletters and receive exclusive discounts and offers on Packt books and eBooks.


http://PacktLib.PacktPub.com

Do you need instant solutions to your IT questions? PacktLib is Packt's online digital book library. Here, you can access, read and search across Packt's entire library of books.

## Why Subscribe?

- Fully searchable across every book published by Packt
- Copy and paste, print and bookmark content
- On demand and accessible via web browser

## Free Access for Packt account holders

If you have an account with Packt at www.PacktPub.com, you can use this to access PacktLib today and view nine entirely free books. Simply use your login credentials for immediate access.

# Instant Updates on New Packt Books

Get notified! Find out when new books are published by following `@PacktEnterprise` on Twitter, or the *Packt Enterprise* Facebook page.

# Preface

Data warehousing as an industry has been around for quite a number of years now. There have been many evolutions in data modeling, storage, and ultimately the vast variety of tools that the business user now has available to help utilize their quickly growing stores of data. As the industry is moving more towards self service business intelligence solutions for the business user, there are also changes in how data is being stored. Amazon Redshift is one of those "game-changing" changes that is not only driving down the total cost, but also driving up the ability to store even more data to enable even better business decisions to be made. This book will not only help you get started in the traditional "how-to" sense, but also provide background and understanding to enable you to make the best use of the data that you already have.

# What this book covers

Chapter 1, *Overview*, takes an in-depth look at what we will be covering in the book, as well as a look at what Redshift provides at the current Amazon pricing levels.

Chapter 2, *Transition to Redshift*, provides the details necessary to start your Redshift cluster. We will begin to look at the tools you will use to connect, as well as the kinds of features that are and are not supported in Redshift.

Chapter 3, *Loading Your Data to Redshift*, will takes you through the steps of creating tables, and the steps necessary to get data loaded into the database.

Chapter 4, *Managing Your Data*, provides you with a good understanding of the day-to-day operation of a Redshift cluster. Everything from backup and recover, to managing user queries with Workload Management is covered here.

Chapter 5, *Querying Data*, gives you the details you need to understand how to monitor the queries you have running, and also helps you to understand explain plans. We will also look at the things you will need to convert your existing queries to Redshift.

Chapter 6, *Best Practices*, will tie together the remaining details about monitoring your Redshift cluster, and provides some guidance on general best practices to get you started in the right direction.

Appendix, *Reference Materials*, will provide you with a point of reference for terms, important commands, and system tables. There is also a consolidated list of links for software, and other utilities discussed in the book.

# What you need for this book

In order to work with the examples, and run your own Amazon Redshift cluster, there are a few things you will need, which are as follows:.

- An Amazon Web Services account with permissions to create and manage Redshift
- Software and drivers (links in the Appendix, *Reference Materials*)
- Client JDBC drivers
- Client ODBC drivers (optional)
- An Amazon S3 file management utility (such as Cloudberry Explorer)
- Query software (such as EMS SQL Manager)
- An Amazon EC2 instance (optional) for the command-line interface

# Who this book is for

This book is intended to provide a practical as well as a technical overview for everyone who is interested in this technology. There is something here for everyone interested in this technology. The CIOs will gain an understanding of what their technical staff is talking about, and the technical implementation personnel will get an in-depth view of the technology and what it will take to implement their own solutions.

# Conventions

In this book, you will find a number of styles of text that distinguish between different kinds of information. Here are some examples of these styles, and an explanation of their meaning.

Code words in text, database table names, folder names, filenames, file extensions, pathnames, dummy URLs, user input, and Twitter handles are shown as follows: "We can include other contexts through the use of the `include` directive."

A block of code is set as follows:

```
CREATE TABLE census_data
   (
       fips                        VARCHAR(10),
       pop_estimate                BIGINT,
       pop_estimate_base           BIGINT,
       pop_estimate_chg            DECIMAL(5, 1),
       pop_total                   BIGINT
...
```

When we wish to draw your attention to a particular part of a code block, the relevant lines or items are set in bold:

```
CREATE TABLE census_data
   (
       fips                        VARCHAR(10),
       pop_estimate                BIGINT,
       pop_estimate_base           BIGINT,
       pop_estimate_chg            DECIMAL(5, 1),
       pop_total                   BIGINT
...
```

Any command-line input or output is written as follows:

```
# cexport AWS_CONFIG_FILE=/home/user/cliconfig.txt
```

**New terms** and **important words** are shown in bold. Words that you see on the screen, in menus or dialog boxes for example, appear in the text like this: "Launch the cluster creation wizard by selecting the **Launch Cluster** option from the Amazon Redshift Management console."

> **Note**

Warnings or important notes appear in a box like this.

## Tip

Tips and tricks appear like this.

# Reader feedback

Feedback from our readers is always welcome. Let us know what you think about this book—what you liked or may have disliked. Reader feedback is important for us to develop titles that you really get the most out of.

To send us general feedback, simply send an e-mail to <feedback@packtpub.com>, and mention the book title via the subject of your message.

If there is a topic that you have expertise in and you are interested in either writing or contributing to a book, see our author guide on www.packtpub.com/authors.

# Customer support

Now that you are the proud owner of a Packt book, we have a number of things to help you to get the most from your purchase.

## Downloading the example code

You can download the example code files for all Packt books you have purchased from your account at http://www.packtpub.com. If you purchased this book elsewhere, you can visit http://www.packtpub.com/support and register to have the files e-mailed directly to you.

## Errata

Although we have taken every care to ensure the accuracy of our content, mistakes do happen. If you find a mistake in one of our books—maybe a mistake in the text or the code—we would be grateful if you would report this to us. By doing so, you can save other readers from frustration and help us improve subsequent versions of this book. If you find any errata, please report them by visiting http://www.packtpub.com/submit-errata, selecting your book, clicking on the **errata submission form** link, and entering the details of your errata. Once your errata are verified, your submission will be accepted and the errata will be uploaded on our website, or added to any list of existing errata, under the Errata section of that title. Any existing errata can be viewed by selecting your title from http://www.packtpub.com/support.

## Piracy

Piracy of copyright material on the Internet is an ongoing problem across all media. At Packt, we take the protection of our copyright and licenses very seriously. If you come across any illegal copies of our works, in any form, on the Internet, please provide us with the location address or website name immediately so that we can pursue a remedy.

Please contact us at <copyright@packtpub.com> with a link to the suspected pirated material.

We appreciate your help in protecting our authors, and our ability to bring you valuable content.

## Questions

You can contact us at <questions@packtpub.com> if you are having a problem with any aspect of the book, and we will do our best to address it.