



Community Experience Distilled

Apache Kafka

Set up Apache Kafka clusters and develop custom message producers and consumers using practical, hands-on examples.

Nishant Garg

[PACKT] open source*
PUBLISHING community experience distilled

Apache Kafka

Set up Apache Kafka clusters and develop custom message producers and consumers using practical, hands-on examples

Nishant Garg



BIRMINGHAM - MUMBAI

Apache Kafka

Copyright © 2013 Packt Publishing

All rights reserved. No part of this book may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, without the prior written permission of the publisher, except in the case of brief quotations embedded in critical articles or reviews.

Every effort has been made in the preparation of this book to ensure the accuracy of the information presented. However, the information contained in this book is sold without warranty, either express or implied. Neither the author, nor Packt Publishing, and its dealers and distributors will be held liable for any damages caused or alleged to be caused directly or indirectly by this book.

Packt Publishing has endeavored to provide trademark information about all of the companies and products mentioned in this book by the appropriate use of capitals. However, Packt Publishing cannot guarantee the accuracy of this information.

First published: October 2013

Production Reference: 1101013

Published by Packt Publishing Ltd.
Livery Place
35 Livery Street
Birmingham B3 2PB, UK..

ISBN 978-1-78216-793-8

www.packtpub.com

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

Credits

Author

Nishant Garg

Project Coordinator

Esha Thakker

Reviewers

Magnus Edenhill

Iuliia Proskurnia

Proofreader

Christopher Smith

Acquisition Editors

Usha Iyer

Julian Ursell

Indexers

Monica Ajmera

Hemangini Bari

Tejal Daruwale

Commissioning Editor

Shaon Basu

Graphics

Abhinash Sahu

Technical Editor

Veena Pagare

Production Coordinator

Kirtee Shingan

Copy Editors

Tanvi Gaitonde

Sayanee Mukherjee

Aditya Nair

Kirti Pai

Alfida Paiva

Adithi Shetty

Cover Work

Kirtee Shingan

About the Author

Nishant Garg is a Technical Architect with more than 13 years' experience in various technologies such as Java Enterprise Edition, Spring, Hibernate, Hadoop, Hive, Flume, Sqoop, Oozie, Spark, Kafka, Storm, Mahout, and Solr/Lucene; NoSQL databases such as MongoDB, CouchDB, HBase and Cassandra, and MPP Databases such as GreenPlum and Vertica.

He has attained his M.S. in Software Systems from Birla Institute of Technology and Science, Pilani, India, and is currently a part of Big Data R&D team in innovation labs at Impetus Infotech Pvt. Ltd.

Nishant has enjoyed working with recognizable names in IT services and financial industries, employing full software lifecycle methodologies such as Agile and SCRUM. He has also undertaken many speaking engagements on Big Data technologies.

I would like to thank my parents (Sh. Vishnu Murti Garg and Smt. Vimla Garg) for their continuous encouragement and motivation throughout my life. I would also like to thank my wife (Himani) and my kids (Nitigya and Darsh) for their never-ending support, which keeps me going.

Finally, I would like to thank Vineet Tyagi – AVP and Head of Innovation Labs, Impetus – and Dr. Vijay – Director of Technology, Innovation Labs, Impetus – for having faith in me and giving me an opportunity to write.

About the Reviewers

Magnus Edenhill is a freelance systems developer living in Stockholm, Sweden, with his family. He specializes in high-performance distributed systems but is also a veteran in embedded systems.

For ten years, Magnus played an instrumental role in the design and implementation of PacketFront's broadband architecture, serving millions of FTTH end customers worldwide. Since 2010, he has been running his own consultancy business with customers ranging from Headweb – northern Europe's largest movie streaming service – to Wikipedia.

Iuliia Proskurnia is a doctoral student at EDIC school of EPFL, specializing in Distributed Computing. Iuliia was awarded the EPFL fellowship to conduct her doctoral research. She is a winner of the Google Anita Borg scholarship and was the Google Ambassador at KTH (2012-2013). She obtained a Masters Diploma in Distributed Computing (2013) from KTH, Stockholm, Sweden, and UPC, Barcelona, Spain. For her Master's thesis, she designed and implemented a unique real-time, low-latency, reliable, and strongly consistent distributed data store for the stock exchange environment at NASDAQ OMX. Previously, she has obtained Master's and Bachelor's Diplomas with honors in Computer Science from the National Technical University of Ukraine KPI. This Master's thesis was about fuzzy portfolio management in previously uncertain conditions. This period was productive for her in terms of publications and conference presentations. During her studies in Ukraine, she obtained several scholarships. During her stay in Kiev, Ukraine, she worked as Financial Analyst at Alfa Bank Ukraine.

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.

Table of Contents

Preface	1
Chapter 1: Introducing Kafka	5
Need for Kafka	7
Few Kafka usages	8
Summary	9
Chapter 2: Installing Kafka	11
Installing Kafka	12
Downloading Kafka	12
Installing the prerequisites	13
Installing Java 1.6 or later	13
Building Kafka	14
Summary	16
Chapter 3: Setting up the Kafka Cluster	17
Single node – single broker cluster	17
Starting the ZooKeeper server	18
Starting the Kafka broker	19
Creating a Kafka topic	20
Starting a producer for sending messages	20
Starting a consumer for consuming messages	22
Single node – multiple broker cluster	23
Starting ZooKeeper	23
Starting the Kafka broker	23
Creating a Kafka topic	24
Starting a producer for sending messages	24
Starting a consumer for consuming messages	25
Multiple node – multiple broker cluster	25
Kafka broker property list	26
Summary	26

Chapter 4: Kafka Design	27
Kafka design fundamentals	28
Message compression in Kafka	29
Cluster mirroring in Kafka	30
Replication in Kafka	31
Summary	32
Chapter 5: Writing Producers	33
The Java producer API	34
Simple Java producer	36
Importing classes	36
Defining properties	36
Building the message and sending it	37
Creating a simple Java producer with message partitioning	38
Importing classes	38
Defining properties	38
Implementing the Partitioner class	39
Building the message and sending it	39
The Kafka producer property list	40
Summary	42
Chapter 6: Writing Consumers	43
Java consumer API	44
High-level consumer API	44
Simple consumer API	46
Simple high-level Java consumer	47
Importing classes	47
Defining properties	47
Reading messages from a topic and printing them	48
Multithreaded consumer for multipartition topics	50
Importing classes	50
Defining properties	50
Reading the message from threads and printing it	51
Kafka consumer property list	54
Summary	55
Chapter 7: Kafka Integrations	57
Kafka integration with Storm	57
Introduction to Storm	58
Integrating Storm	59
Kafka integration with Hadoop	60
Introduction to Hadoop	60
Integrating Hadoop	62