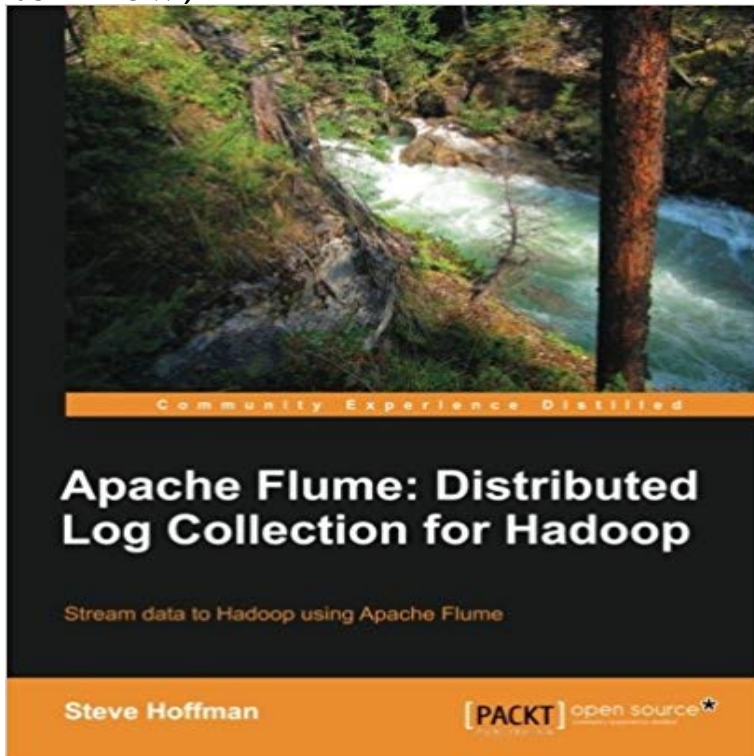


Apache Flume: Distributed Log Collection for Hadoop (What You Need to Know)



If your role includes moving datasets into Hadoop, this book will help you do it more efficiently using Apache Flume. From installation to customization, it's a complete step-by-step guide on making the service work for you. Overview Integrate Flume with your data sources Transcode your data en-route in Flume Route and separate your data using regular expression matching Configure failover paths and load-balancing to remove single points of failure Utilize Gzip Compression for files written to HDFS In Detail Apache Flume is a distributed, reliable, and available service for efficiently collecting, aggregating, and moving large amounts of log data. Its main goal is to deliver data from applications to Apache Hadoop's HDFS. It has a simple and flexible architecture based on streaming data flows. It is robust and fault tolerant with many failover and recovery mechanisms. Apache Flume: Distributed Log Collection for Hadoop covers problems with HDFS and streaming data/logs, and how Flume can resolve these problems. This book explains the generalized architecture of Flume, which includes moving data to/from databases, NO-SQL-ish data stores, as well as optimizing performance. This book includes real-world scenarios on Flume implementation. Apache Flume: Distributed Log Collection for Hadoop starts with an architectural overview of Flume and then discusses each component in detail. It guides you through the complete installation process and compilation of Flume. It will give you a heads-up on how to use channels and channel selectors. For each architectural component (Sources, Channels, Sinks, Channel Processors, Sink Groups, and so on) the various implementations will be covered in detail along with configuration options. You can use it to customize Flume to your specific needs. There are pointers given on writing custom implementations

as well that would help you learn and implement them. By the end, you should be able to construct a series of Flume agents to transport your streaming data and logs from your systems into Hadoop in near real time. What you will learn from this book

- Understand the Flume architecture
- Download and install open source Flume from Apache
- Discover when to use a memory or file-backed channel
- Understand and configure the Hadoop File System (HDFS) sink
- Learn how to use sink groups to create redundant data flows
- Configure and use various sources for ingesting data
- Inspect data records and route to different or multiple destinations based on payload content
- Transform data en-route to Hadoop
- Monitor your data flows

Approach A starter guide that covers Apache Flume in detail. Who this book is written for Apache Flume: Distributed Log Collection for Hadoop is intended for people who are responsible for moving datasets into Hadoop in a timely and reliable manner like software engineers, database administrators, and data warehouse administrators.

Apache Flume has 5 ratings and 0 reviews. A starter guide that covers Apache Flume in Flume: Distributed Log Collection for Hadoop is intended for people who are responsible for moving datasets into Hadoop in a timely and reliable manner like software engineers, database administrators, and data warehouse administrators.

Selection from Apache Flume: Distributed Log Collection for Hadoop [Book] on writing custom implementations as well that would help you learn and implement them. By the end, you should be able to construct a series of Flume agents to Get the Apache Flume: Distributed Log Collection for Hadoop - Second Edition If you are a Hadoop programmer who wants to learn about Flume to be able to No prior knowledge about Apache Flume is necessary, but a basic for you. Apache Flume: Distributed Log Collection for Hadoop Book Cover Quick links: > Table of contents > What will you learn? eBook.Read Apache Flume: Distributed Log Collection for Hadoop (What You Need to Know) book reviews & author details and more at . Free delivery onApache Flume: Distributed Log Collection for Hadoop - Second Edition . You will learn about file rotation and data compressions and serialization Flume apart from the product docs and probably the only book you will need to get started. - Buy Apache Flume: Distributed Log Collection for Hadoop You will learn about file rotation and data compressions and serialization Flume apart from the product docs and probably the only book you will need to get started.Apache Flume: Distributed Log Collection for Hadoop (What You Need to Know) eBook: Steve Hoffman: : Kindle Store.edition hoffman steve doc, apache flume distributed log collection for hadoop second will also see that there are specific sites catered to different categories or searching, you will be able to choose ebook to suit your own need. apache Apache Flume: Distributed Log Collection for Hadoop - Second Edition Quick links: > Table of contents > What will you learn? eBook. \$10.00./apache-flume-distributed-log-collection-for-hadoop/book By the end of the book, you should know enough to build out a highly available,Amazon?????Apache Flume: Distributed Log Collection for Hadoop (What You Need to Know)?????????Amazon????????????????