

---

# Apache Nifi Overview Linux Foundation

---

Apache NiFi Introduction Learning the Basics of Apache NiFi for IoT - Timothy J Spann, Cloudera Apache NiFi 101 | How to get started with Apache NiFi Workshop: Log Processing with Apache NiFi and Apache Kafka - Sophia Izokun, IBM Apache Nifi Crash Course Chapter 1 - Apache NiFi Core Concepts and Architecture GLT #1 - What is a table format? (like Apache Iceberg, Delta Lake and Apache Hudi) IoT with Apache MXNet and Apache NiFi and MiniFi Airflow Tutorial for Beginners - Full Course in 2 Hours 2022 NiFi Flows Demo The Harsh Reality of Being a Data Analyst Apache NiFi REST API Basic Access and Usage NiFi vs Kafka | Differences Between NiFi and Kafka Tutorial: Building a Data Lakehouse with Apache Iceberg, Spark, Dremio, Nessie \u0026 Minio What is Apache NiFi | Apache NiFi Tutorial | Online Apache NiFi Training Apache Nifi Basics 2, Transformations Apache Nifi Basics Apache NiFi Basics for beginners Apache NiFi Tutorial - Complete Guide (Part 1) - Course Introduction |

#ApacheNiFi Secure IoT Command, Control, and Exfil using Apache NiFi and Apache MiNiFi - Andy LoPresto, Cloudera Apache NiFi Tutorial - Complete Guide (Part 4) - What is Apache NiFi? | #ApacheNiFi Apache Nifi | Apache Nifi Tutorial | Apache Nifi Crash Course | Intellipaat Apache NiFi Important Concepts Explained Plasma 6.1: the BEST LINUX DESKTOP (in my opinion)  
Enterprise Data Warehouse Optimization with Hadoop on IBM Power Systems Servers  
Data Lake for Enterprises  
Open Source Democracy  
Building Big Data Applications  
Microservices for the Enterprise  
Hadoop 2 Quick-Start Guide  
Code Like a Pro in C#  
Spark for Data Science  
Practical Data Science with Hadoop and Spark  
Stream Processing with Apache Flink  
The Architecture of Open Source Applications  
Agile Data Science  
Kafka: The Definitive Guide  
Kafka Streams in Action  
Disruptive Analytics

Data Engineering with Python  
Scala Programming for Big Data Analytics  
Open Reference Architecture for Security and Privacy

*Apache Nifi*                      *OMB No.*  
*Overview Linux 4207385916248*  
*Foundation*                      *edited by*

---

**MAXIMILLIAN  
JANIYAH**

---

Enterprise Data  
Warehouse Optimization  
with Hadoop on IBM  
Power Systems Servers  
Demos

Data-driven insights are a key competitive advantage for any industry today, but deriving insights from raw data can still take days or

weeks. Most organizations can't scale data science teams fast enough to keep up with the growing amounts of data to transform. What's the answer? Self-service data. With this practical book, data engineers, data scientists, and team managers will learn how to build a self-service data science platform that helps anyone in your organization extract insights from data.

Sandeep Uttamchandani provides a scorecard to track and address bottlenecks that slow down time to insight across data discovery, transformation, processing, and production. This book bridges the gap between data scientists bottlenecked by engineering realities and data engineers unclear about ways to make self-service work. Build a self-

service portal to support data discovery, quality, lineage, and governance  
 Select the best approach for each self-service capability using open source cloud technologies  
 Tailor self-service for the people, processes, and technology maturity of your data platform  
 Implement capabilities to democratize data and reduce time to insight  
 Scale your self-service portal to support a large number of users within your organization

## DATA LAKE FOR ENTERPRISES

Apress  
 Political structures need to change. They will emerge from people acting and communicating in the present, not talking about a fictional future ... (from cover).  
[Open Source Democracy](#)  
 Simon and Schuster  
 Due to the continuously stream of security breaches two security architects in the Netherlands started a project to harvest good

practices for better and faster creating architecture and privacy solution designs. This project resulted in a reference architecture that is aimed to help all security architects and designers worldwide. All kinds of topics that help creating a security or privacy solution architecture are outlined, such as: security and privacy principles, common attack vectors, threat models while in-depth guidelines are also given to evaluate the use of Open Source security

and privacy application in various use cases.

*Building Big Data Applications* Packt Publishing Ltd

The Complete Guide to

Data Science with Hadoop—For Technical Professionals,

Businesspeople, and Students Demand is

soaring for professionals who can solve real data

science problems with Hadoop and Spark.

Practical Data Science with Hadoop® and Spark

is your complete guide to doing just that. Drawing

on immense experience

with Hadoop and big data, three leading experts

bring together everything you need: high-level

concepts, deep-dive techniques, real-world use

cases, practical applications, and hands-

on tutorials. The authors introduce the essentials of

data science and the modern Hadoop

ecosystem, explaining how Hadoop and Spark

have evolved into an effective platform for

solving data science problems at scale. In

addition to comprehensive

application coverage, the authors also provide

useful guidance on the important steps of data

ingestion, data munging, and visualization. Once

the groundwork is in place, the authors focus

on specific applications, including machine

learning, predictive modeling for sentiment

analysis, clustering for document analysis,

anomaly detection, and natural language

processing (NLP). This guide provides a strong

technical foundation for those who want to do

practical data science, and also presents business-driven guidance on how to apply Hadoop and Spark to optimize ROI of data science initiatives. Learn What data science is, how it has evolved, and how to plan a data science career How data volume, variety, and velocity shape data science use cases Hadoop and its ecosystem, including HDFS, MapReduce, YARN, and Spark Data importation with Hive and Spark Data quality, preprocessing, preparation, and modeling

Visualization: surfacing insights from huge data sets Machine learning: classification, regression, clustering, and anomaly detection Algorithms and Hadoop tools for predictive modeling Cluster analysis and similarity functions Large-scale anomaly detection NLP: applying data science to human language Microservices for the Enterprise Packt Publishing Ltd Get expert guidance on architecting end-to-end data management

solutions with Apache Hadoop. While many sources explain how to use various components in the Hadoop ecosystem, this practical book takes you through architectural considerations necessary to tie those components together into a complete tailored application, based on your particular use case. To reinforce those lessons, the book's second section provides detailed examples of architectures used in some of the most commonly found Hadoop applications. Whether

you're designing a new Hadoop application, or planning to integrate Hadoop into your existing data infrastructure, Hadoop Application Architectures will skillfully guide you through the process. This book covers: Factors to consider when using Hadoop to store and model data Best practices for moving data in and out of the system Data processing frameworks, including MapReduce, Spark, and Hive Common Hadoop processing patterns, such as removing duplicate

records and using windowing analytics Giraph, GraphX, and other tools for large graph processing on Hadoop Using workflow orchestration and scheduling tools such as Apache Oozie Near-real-time stream processing with Apache Storm, Apache Spark Streaming, and Apache Flume Architecture examples for clickstream analysis, fraud detection, and data warehousing *Hadoop 2 Quick-Start Guide* IBM Redbooks Every enterprise

application creates data, whether it's log messages, metrics, user activity, outgoing messages, or something else. And how to move all of this data becomes nearly as important as the data itself. If you're an application architect, developer, or production engineer new to Apache Kafka, this practical guide shows you how to use this open source streaming platform to handle real-time data feeds. Engineers from Confluent and LinkedIn who are responsible for developing

Kafka explain how to deploy production Kafka clusters, write reliable event-driven microservices, and build scalable stream-processing applications with this platform. Through detailed examples, you'll learn Kafka's design principles, reliability guarantees, key APIs, and architecture details, including the replication protocol, the controller, and the storage layer. Understand publish-subscribe messaging and how it fits in the big data ecosystem.

Explore Kafka producers and consumers for writing and reading messages Understand Kafka patterns and use-case requirements to ensure reliable data delivery Get best practices for building data pipelines and applications with Kafka Manage Kafka in production, and learn to perform monitoring, tuning, and maintenance tasks Learn the most critical metrics among Kafka's operational measurements Explore how Kafka's stream delivery capabilities make

it a perfect source for stream processing systems  
**Code Like a Pro in C#**  
"O'Reilly Media, Inc."  
Get started with Apache Flink, the open source framework that powers some of the world's largest stream processing applications. With this practical book, you'll explore the fundamental concepts of parallel stream processing and discover how this technology differs from traditional batch data processing. Longtime Apache Flink committers



Fabian Hueske and Vasia Kalavri show you how to implement scalable streaming applications with Flink's DataStream API and continuously run and maintain these applications in operational environments. Stream processing is ideal for many use cases, including low-latency ETL, streaming analytics, and real-time dashboards as well as fraud detection, anomaly detection, and alerting. You can process continuous data of any kind, including user interactions, financial

transactions, and IoT data, as soon as you generate them. Learn concepts and challenges of distributed stateful stream processing Explore Flink's system architecture, including its event-time processing mode and fault-tolerance model Understand the fundamentals and building blocks of the DataStream API, including its time-based and stateful operators Read data from and write data to external systems with exactly-once consistency Deploy and configure

Flink clusters Operate continuously running streaming applications

## **SPARK FOR DATA SCIENCE**

Springer Nature  
This book teaches you how to build and maintain effective data pipelines. You'll explore the most common usage patterns, including aggregating multiple data sources, connecting to and from data lakes, and cloud deployment. --

## **PRACTICAL DATA**

## SCIENCE WITH HADOOP AND SPARK

"O'Reilly Media, Inc."  
Mining big data requires a deep investment in people and time. How can you be sure you're building the right models? With this hands-on book, you'll learn a flexible toolset and methodology for building effective analytics applications with Hadoop. Using lightweight tools such as Python, Apache Pig, and the D3.js library, your team will create an agile environment for exploring

data, starting with an example application to mine your own email inboxes. You'll learn an iterative approach that enables you to quickly change the kind of analysis you're doing, depending on what the data is telling you. All example code in this book is available as working Heroku apps. Create analytics applications by using the agile big data development methodology Build value from your data in a series of agile sprints, using the data-value stack Gain

insight by using several data structures to extract multiple features from a single dataset Visualize data with charts, and expose different aspects through interactive reports Use historical data to predict the future, and translate predictions into action Get feedback from users after each sprint to keep your project on track [Stream Processing with Apache Flink](#) O'Reilly Media Analyze vast amounts of data in record time using Apache Spark with Databricks in the Cloud.

Learn the fundamentals, and more, of running analytics on large clusters in Azure and AWS, using Apache Spark with Databricks on top.

Discover how to squeeze the most value out of your data at a mere fraction of what classical analytics solutions cost, while at the same time getting the results you need, incrementally faster. This book explains how the confluence of these pivotal technologies gives you enormous power, and cheaply, when it comes to huge datasets. You will

begin by learning how cloud infrastructure makes it possible to scale your code to large amounts of processing units, without having to pay for the machinery in advance. From there you will learn how Apache Spark, an open source framework, can enable all those CPUs for data analytics use. Finally, you will see how services such as Databricks provide the power of Apache Spark, without you having to know anything about configuring hardware or software. By removing the

need for expensive experts and hardware, your resources can instead be allocated to actually finding business value in the data. This book guides you through some advanced topics such as analytics in the cloud, data lakes, data ingestion, architecture, machine learning, and tools, including Apache Spark, Apache Hadoop, Apache Hive, Python, and SQL. Valuable exercises help reinforce what you have learned. What You Will Learn Discover the value of big data analytics

that leverage the power of the cloud. Get started with Databricks using SQL and Python in either Microsoft Azure or AWS. Understand the underlying technology, and how the cloud and Apache Spark fit into the bigger picture. See how these tools are used in the real world. Run basic analytics, including machine learning, on billions of rows at a fraction of a cost or free. Who This Book Is For: Data engineers, data scientists, and cloud architects who want or need to run

advanced analytics in the cloud. It is assumed that the reader has data experience, but perhaps minimal exposure to Apache Spark and Azure Databricks. The book is also recommended for people who want to get started in the analytics field, as it provides a strong foundation.

### **THE ARCHITECTURE OF OPEN SOURCE APPLICATIONS**

"O'Reilly Media, Inc." Data is bigger, arrives faster, and comes in a variety of formats—and it

all needs to be processed at scale for analytics or machine learning. But how can you process such varied workloads efficiently? Enter Apache Spark. Updated to include Spark 3.0, this second edition shows data engineers and data scientists why structure and unification in Spark matters. Specifically, this book explains how to perform simple and complex data analytics and employ machine learning algorithms. Through step-by-step walk-throughs, code

snippets, and notebooks, you'll be able to: Learn Python, SQL, Scala, or Java high-level Structured APIs Understand Spark operations and SQL Engine Inspect, tune, and debug Spark operations with Spark configurations and Spark UI Connect to data sources: JSON, Parquet, CSV, Avro, ORC, Hive, S3, or Kafka Perform analytics on batch and streaming data using Structured Streaming Build reliable data pipelines with open source Delta Lake and Spark Develop machine

learning pipelines with MLlib and productionize models using MLflow

## **AGILE DATA SCIENCE**

O'Reilly Media  
Why a book about logs? That's easy: the humble log is an abstraction that lies at the heart of many systems, from NoSQL databases to cryptocurrencies. Even though most engineers don't think much about them, this short book shows you why logs are worthy of your attention. Based on his popular blog posts, LinkedIn principal

engineer Jay Kreps shows you how logs work in distributed systems, and then delivers practical applications of these concepts in a variety of common uses—data integration, enterprise architecture, real-time stream processing, data system design, and abstract computing models. Go ahead and take the plunge with logs; you're going love them. Learn how logs are used for programmatic access in databases and distributed systems Discover solutions to the

huge data integration problem when more data of more varieties meet more systems Understand why logs are at the heart of real-time stream processing Learn the role of a log in the internals of online data systems Explore how Jay Kreps applies these ideas to his own work on data infrastructure systems at LinkedIn

## **KAFKA: THE DEFINITIVE GUIDE**

Flow Architectures  
Learn all you need to know about seven key

innovations disrupting business analytics today. These innovations—the open source business model, cloud analytics, the Hadoop ecosystem, Spark and in-memory analytics, streaming analytics, Deep Learning, and self-service analytics—are radically changing how businesses use data for competitive advantage. Taken together, they are disrupting the business analytics value chain, creating new opportunities. Enterprises who seize the opportunity

will thrive and prosper, while others struggle and decline: disrupt or be disrupted. Disruptive Business Analytics provides strategies to profit from disruption. It shows you how to organize for insight, build and provision an open source stack, how to practice lean data warehousing, and how to assimilate disruptive innovations into an organization. Through a short history of business analytics and a detailed survey of products and services, analytics

authority Thomas W. Dinsmore provides a practical explanation of the most compelling innovations available today. What You'll Learn Discover how the open source business model works and how to make it work for you See how cloud computing completely changes the economics of analytics Harness the power of Hadoop and its ecosystem Find out why Apache Spark is everywhere Discover the potential of streaming and real-time analytics Learn what Deep

Learning can do and why it matters See how self-service analytics can change the way organizations do business Who This Book Is For Corporate actors at all levels of responsibility for analytics: analysts, CIOs, CTOs, strategic decision makers, managers, systems architects, technical marketers, product developers, IT personnel, and consultants.

### **KAFKA STREAMS IN ACTION**

"O'Reilly Media, Inc."

Get Started Fast with Apache Hadoop® 2, YARN, and Today's Hadoop Ecosystem With Hadoop 2.x and YARN, Hadoop moves beyond MapReduce to become practical for virtually any type of data processing. Hadoop 2.x and the Data Lake concept represent a radical shift away from conventional approaches to data usage and storage. Hadoop 2.x installations offer unmatched scalability and breakthrough extensibility that supports new and existing Big Data analytics

processing methods and models. Hadoop® 2 Quick-Start Guide is the first easy, accessible guide to Apache Hadoop 2.x, YARN, and the modern Hadoop ecosystem. Building on his unsurpassed experience teaching Hadoop and Big Data, author Douglas Eadline covers all the basics you need to know to install and use Hadoop 2 on personal computers or servers, and to navigate the powerful technologies that complement it. Eadline concisely

introduces and explains every key Hadoop 2 concept, tool, and service, illustrating each with a simple “beginning-to-end” example and identifying trustworthy, up-to-date resources for learning more. This guide is ideal if you want to learn about Hadoop 2 without getting mired in technical details. Douglas Eadline will bring you up to speed quickly, whether you’re a user, admin, devops specialist, programmer, architect, analyst, or data scientist. Coverage Includes Understanding what

Hadoop 2 and YARN do, and how they improve on Hadoop 1 with MapReduce Understanding Hadoop-based Data Lakes versus RDBMS Data Warehouses Installing Hadoop 2 and core services on Linux machines, virtualized sandboxes, or clusters Exploring the Hadoop Distributed File System (HDFS) Understanding the essentials of MapReduce and YARN application programming Simplifying programming and data movement with Apache Pig, Hive, Sqoop, Flume,



Oozie, and HBase  
Observing application progress, controlling jobs, and managing workflows  
Managing Hadoop efficiently with Apache Ambari-including recipes for HDFS to NFSv3 gateway, HDFS snapshots, and YARN configuration  
Learning basic Hadoop 2 troubleshooting, and installing Apache Hue and Apache Spark  
Disruptive Analytics  
Simon and Schuster  
Data warehouses were developed for many good reasons, such as providing quick query and

reporting for business operations, and business performance. However, over the years, due to the explosion of applications and data volume, many existing data warehouses have become difficult to manage. Extract, Transform, and Load (ETL) processes are taking longer, missing their allocated batch windows. In addition, data types that are required for business analysis have expanded from structured data to unstructured data. The Apache open source Hadoop platform provides

a great alternative for solving these problems. IBM® has committed to open source since the early years of open Linux. IBM and Hortonworks together are committed to Apache open source software more than any other company. IBM Power Systems™ servers are built with open technologies and are designed for mission-critical data applications. Power Systems servers use technology from the OpenPOWER Foundation, an open technology infrastructure that uses

the IBM POWER® architecture to help meet the evolving needs of big data applications. The combination of Power Systems with Hortonworks Data Platform (HDP) provides users with a highly efficient platform that provides leadership performance for big data workloads such as Hadoop and Spark. This IBM Redpaper™ publication provides details about Enterprise Data Warehouse (EDW) optimization with Hadoop on Power Systems. Many people know Power

Systems from the IBM AIX® platform, but might not be familiar with IBM PowerLinux™, so part of this paper provides a Power Systems overview. A quick introduction to Hadoop is provided for those not familiar with the topic. Details of HDP on Power Reference architecture are included that will help both software architects and infrastructure architects understand the design. In the optimization chapter, we describe various topics: traditional EDW offload, sizing guidelines,

performance tuning, IBM Elastic Storage™ Server (ESS) for data-intensive workload, IBM Big SQL as the common structured query language (SQL) engine for Hadoop platform, and tools that are available on Power Systems that are related to EDW optimization. We also dedicate some pages to the analytics components (IBM Data Science Experience (IBM DSX) and IBM Spectrum™ Conductor for Spark workload) for the Hadoop infrastructure. **Data Engineering with**

**Python** Addison-Wesley Enterprise Integration Patterns provides an invaluable catalog of sixty-five patterns, with real-world solutions that demonstrate the formidable of messaging and help you to design effective messaging solutions for your enterprise. The authors also include examples covering a variety of different integration technologies, such as JMS, MSMQ, TIBCO ActiveEnterprise, Microsoft BizTalk, SOAP, and XSL. A case study

describing a bond trading system illustrates the patterns in practice, and the book offers a look at emerging standards, as well as insights into what the future of enterprise integration might hold. This book provides a consistent vocabulary and visual notation framework to describe large-scale integration solutions across many technologies. It also explores in detail the advantages and limitations of asynchronous messaging architectures. The authors

present practical advice on designing code that connects an application to a messaging system, and provide extensive information to help you determine when to send a message, how to route it to the proper destination, and how to monitor the health of a messaging system. If you want to know how to manage, monitor, and maintain a messaging system once it is in use, get this book. [Scala Programming for Big Data Analytics](#) "O'Reilly Media, Inc." Clear your doubts about

Business Intelligence and start your new journey

**KEY FEATURES**

- Includes successful methods and innovative ideas to achieve success with BI.
- Vendor-neutral, unbiased, and based on experience.
- Highlights practical challenges in BI journeys.
- Covers financial aspects along with technical aspects.
- Showcases multiple BI organization models and the structure of BI teams.

**DESCRIPTION** The book demystifies misconceptions and misinformation about BI.

It provides clarity to almost everything related to BI in a simplified and unbiased way. It covers topics right from the definition of BI, terms used in the BI definition, coinage of BI, details of the different main uses of BI, processes that support the main uses, side benefits, and the level of importance of BI, various types of BI based on various parameters, main phases in the BI journey and the challenges faced in each of the phases in the BI journey. It clarifies myths about self-service

BI and real-time BI. The book covers the structure of a typical internal BI team, BI organizational models, and the main roles in BI. It also clarifies the doubts around roles in BI. It explores the different components that add to the cost of BI and explains how to calculate the total cost of the ownership of BI and ROI for BI. It covers several ideas, including unconventional ideas to achieve BI success and also learn about IBI. It explains the different types of BI architectures,

commonly used technologies, tools, and concepts in BI and provides clarity about the boundary of BI w.r.t technologies, tools, and concepts. The book helps you lay a very strong foundation and provides the right perspective about BI. It enables you to start or restart your journey with BI. WHAT YOU WILL LEARN ● Builds a strong conceptual foundation in BI. ● Gives the right perspective and clarity on BI uses, challenges, and architectures. ● Enables

you to make the right decisions on the BI structure, organization model, and budget. ● Explains which type of BI solution is required for your business. ● Applies successful BI ideas. WHO THIS BOOK IS FOR This book is a must-read for business managers, BI aspirants, CxOs, and all those who want to drive the business value with data-driven insights. TABLE OF CONTENTS 1. What is Business Intelligence? 2. Why do Businesses need BI? 3. Types of Business

Intelligence 4. Challenges in Business Intelligence 5. Roles in Business Intelligence 6. Financials of Business Intelligence 7. Ideas for Success with BI 8. Introduction to IBI 9. BI Architectures 10. Demystify Tech, Tools, and Concepts in BI *Open Reference Architecture for Security and Privacy* Apress As big data becomes more ubiquitous, businesses are wondering how they can best leverage it to gain insight into their most important business questions. Using

machine learning (ML) and deep learning (DL) in big data environments can identify historical patterns and build artificial intelligence (AI) models that can help businesses to improve customer experience, add services and offerings, identify new revenue streams or lines of business (LOBs), and optimize business or manufacturing operations. The power of AI for predictive analytics is being harnessed across all industries, so it is important that businesses

familiarize themselves with all of the tools and techniques that are available for integration with their data lake environments. In this IBM® Redbooks® publication, we cover the best practices for deploying and integrating some of the best AI solutions on the market, including: IBM Watson Machine Learning Accelerator (see note for product naming) IBM Watson Studio Local IBM Power Systems™ IBM Spectrum™ Scale IBM Data Science Experience

(IBM DSX) IBM Elastic Storage™ Server Hortonworks Data Platform (HDP) Hortonworks DataFlow (HDF) H2O Driverless AI We map out all the integrations that are possible with our different AI solutions and how they can integrate with your existing or new data lake. We also walk you through some of our client use cases and show you how some of the industry leaders are using Hortonworks, IBM PowerAI, and IBM Watson Studio Local to drive

decision making. We also advise you on your deployment options, when to use a GPU, and why you should use the IBM Elastic Storage Server (IBM ESS) to improve storage management. Lastly, we describe how to integrate IBM Watson Machine Learning Accelerator and Hortonworks with or without IBM Watson Studio Local, how to access real-time data, and security. Note: IBM Watson Machine Learning Accelerator is the new product name for IBM

PowerAI Enterprise. Note: Hortonworks merged with Cloudera in January 2019. The new company is called Cloudera. References to Hortonworks as a business entity in this publication are now referring to the merged company. Product names beginning with Hortonworks continue to be marketed and sold under their original names. *Data Pipelines with Apache Airflow* Packt Publishing Ltd Leverage Kubernetes for

the rapid adoption of emerging technologies. Kubernetes is the future of enterprise platform development and has become the most popular, and often considered the most robust, container orchestration system available today. This book focuses on platforming technologies that power the Internet of Things, Blockchain, Machine Learning, and the many layers of data and application management supporting them. Advanced Platform Development with

Kubernetes takes you through the process of building platforms with these in-demand capabilities. You'll progress through the development of Serverless, CICD integration, data processing pipelines, event queues, distributed query engines, modern data warehouses, data lakes, distributed object storage, indexing and analytics, data routing and transformation, query engines, and data science/machine learning environments. You'll also

see how to implement and tie together numerous essential and trending technologies including: Kafka, NiFi, Airflow, Hive, Keycloak, Cassandra, MySQL, Zookeeper, Mosquitto, Elasticsearch, Logstash, Kibana, Presto, Mino, OpenFaaS, and Ethereum. The book uses Golang and Python to demonstrate the development integration of custom container and Serverless functions, including interaction with the Kubernetes API. The exercises throughout

teach Kubernetes through the lens of platform development, expressing the power and flexibility of Kubernetes with clear and pragmatic examples. Discover why Kubernetes is an excellent choice for any individual or organization looking to embark on developing a successful data and application platform. What You'll Learn Configure and install Kubernetes and k3s on vendor-neutral platforms, including generic virtual machines and bare metal Implement an integrated



development toolchain for continuous integration and deployment Use data pipelines with MQTT, NiFi, Logstash, Kafka and Elasticsearch Install a serverless platform with OpenFaaS Explore blockchain network capabilities with Ethereum Support a multi-tenant data science platform and web IDE with JupyterHub, MLflow and Seldon Core Build a hybrid cluster, securely bridging on-premise and cloud-based Kubernetes nodes Who This Book Is For System and software architects,

full-stack developers, programmers, and DevOps engineers with some experience building and using containers. This book also targets readers who have started with Kubernetes and need to progress from a basic understanding of the technology and "Hello World" example to more productive, career-building projects. *I Heart Logs* BPB Publications If you've been asked to maintain large and complex Hadoop clusters, this book is a must.

Demand for operations-specific material has skyrocketed now that Hadoop is becoming the de facto standard for truly large-scale data processing in the data center. Eric Sammer, Principal Solution Architect at Cloudera, shows you the particulars of running Hadoop in production, from planning, installing, and configuring the system to providing ongoing maintenance. Rather than run through all possible scenarios, this pragmatic operations guide calls out what

works, as demonstrated in critical deployments. Get a high-level overview of HDFS and MapReduce: why they exist and how they work Plan a Hadoop deployment, from hardware and OS selection to network

requirements Learn setup and configuration details with a list of critical properties Manage resources by sharing a cluster across multiple groups Get a runbook of the most common cluster

maintenance tasks Monitor Hadoop clusters—and learn troubleshooting with the help of real-world war stories Use basic tools and techniques to handle backup and catastrophic failure

Related with Apache Nifi Overview Linux Foundation:

[© Apache Nifi Overview Linux Foundation Maintenance Technician Test Questions And Answers](#)

[© Apache Nifi Overview Linux Foundation Makeup Exam Or Make Up Exam](#)

[© Apache Nifi Overview Linux Foundation Mafia 2 Definitive Edition Trophy Guide](#)