1. Record Nr. UNINA9910464940403321

Autore	Goetz P. Taylor
Titolo	Storm blueprints : patterns for distributed real-time computation : use Storm design patterns to perform distributed, real-time big data processing, and analytics for real-world use cases / / P. Taylor Goetz , Brian O'Neill ; cover image by Prashant Timappa Shetty
Pubbl/distr/stampa	Birmingham, England : , : Packt Publishing, , 2014 ©2014
ISBN	1-78216-830-3
Edizione	[1st edition]
Descrizione fisica	1 online resource (336 p.)
Collana	Community Experience Distilled
Disciplina	005.73
Soggetti	Data structures (Computer science) Real-time data processing Electronic books.
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Note generali	Includes index.
Nota di contenuto	Cover; Copyright; Credits; About the Authors; About the Reviewers; www.PacktPub.com; Table of Contents; Preface; Chapter 1: Distributed Word Count; Introducing elements of a Storm topology - streams, spouts, and bolts; Streams; Spouts; Bolts; Introducing the word count topology data flow; Sentence spout; Introducing the split sentence bolt; Introducing the word count bolt; Introducing the report bolt; Implementing the word count topology; Setting up a development environment; Implementing the sentence spout; Implementing the split sentence bolt; Implementing the word count bolt Implementing the report boltImplementing the word count topology; Introducing parallelism in Storm; WordCountTopology parallelism; Adding workers to a topology; Configuring executors and tasks; Understanding stream groupings; Guaranteed processing; Reliability in spouts; Reliability in bolts; Reliable word count; Summary; Chapter 2: Configuring Storm Clusters; Introducing the anatomy of a Storm cluster; Understanding the nimbus daemon; Working with the supervisor daemon; Introducing Apache ZooKeeper; Working with Storm's DRPC server; Introducing the Storm UI

Introducing the Storm technology stackJava and Clojure; Python; Installing Storm on Linux; Installing the base operating system; Installing Java; ZooKeeper installation; Storm installation; Running the Storm daemons; Configuring Storm; Mandatory settings; Optional settings; The Storm executable; Setting up the Storm executable on a workstation; The daemon commands; Nimbus; Supervisor; UI; DRPC; The management commands; Jar; Kill; Deactivate; Activate; Rebalance; Remoteconfvalue; Local debug/development commands; REPL; Classpath; Localconfvalue; Submitting topologies to a Storm cluster Automating the cluster configurationA rapid introduction to Puppet; Puppet manifests; Puppet classes and modules; Puppet templates; Managing environments with Puppet Hiera; Introducing Hiera; Summary; Chapter 3: Trident Topologies and Sensor Data; Examining our use case; Introducing Trident topologies; Introducing Trident spouts; Introducing Trident operations - filters and functions; Introducing Trident filters; Introducing Trident functions; Introducing Trident aggregators - Combiners and Reducers; CombinerAggregator; ReducerAggregator; Aggregator; Introducing the Trident state
The Repeat Transactional stateThe Opaque state; Executing the topology; Summary; Chapter 4: Real-time Trend Analysis; Use case; Architecture; The source application; The logback Kafka appender; Apache Kafka; Kafka spout; The XMPP server; Installing the required software; Installing Kafka; Installing OpenFire; Introducing the sample application; Sending log messages to Kafka; Introducing the log analysis topology; Kafka spout; The JSON project function; Calculating a moving average; Adding a sliding window; Implementing the moving average function; Filtering on thresholds
Sending notifications with XMPP

Sommario/riassunto

A blueprints book with 10 different projects built in 10 different chapters which demonstrate the various use cases of storm for both beginner and intermediate users, grounded in real-world example applications.Although the book focuses primarily on Java development with Storm, the patterns are more broadly applicable and the tips, techniques, and approaches described in the book apply to architects, developers, and operations.Additionally, the book should provoke and inspire applications of distributed computing to other industries and domains. Hadoop enthusiasts will also find this book a go