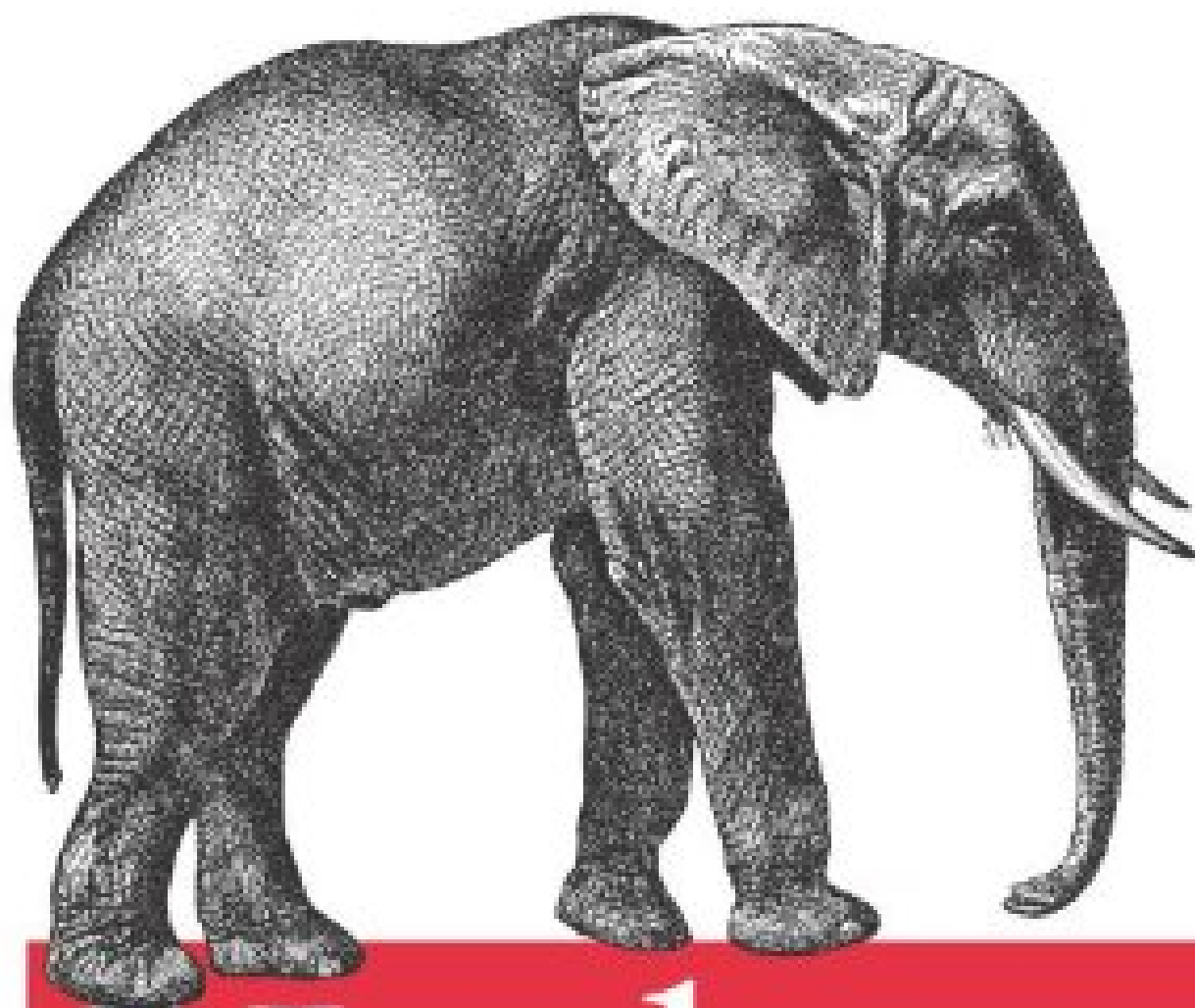Storage and Analysis at Internet Scale

3rd Edition

Revised & Updated

Hadoop

The Definitive Guide

Tom White

Resumo de Hadoop: The Definitive Guide

Ready to unlock the power of your data? With this comprehensive guide, you'll learn how to build and maintain reliable, scalable, distributed systems with Apache Hadoop. This book is ideal for programmers looking to analyze datasets of any size, and for administrators who want to set up and run Hadoop clusters.You'll find illuminating case studies that demonstrate how Hadoop is used to solve specific problems.

This third edition covers recent changes to Hadoop, including material on the new MapReduce API, as well as MapReduce 2 and its more flexible execution model (YARN).Store large datasets with the Hadoop Distributed File System (HDFS)Run distributed computations with MapReduceUse Hadoop's data and I/O building blocks for compression, data integrity, serialization (including Avro), and persistenceDiscover common pitfalls and advanced features for writing real-world MapReduce programsDesign, build, and administer a dedicated Hadoop cluster—or run Hadoop in the cloudLoad data from relational databases into HDFS, using SqoopPerform large-scale data processing with the Pig query languageAnalyze datasets with Hive, Hadoop's data warehousing systemTake advantage of HBase for structured and semi-structured data, and ZooKeeper for building distributed systems

Acesse aqui a versão completa deste livro