Hadoop: the definitive guide

White, Tom, author

Deskripsi Lengkap: https://lib.ui.ac.id/detail?id=20464942&lokasi=lokal

----

Abstrak

Get ready to unlock the power of your data. With the fourth edition of 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.Using Hadoop 2 exclusively, author Tom White presents new chapters on YARN and several Hadoop-related projects such as Parquet, Flume, Crunch, and Spark. You'll learn about recent changes to Hadoop, and explore new case studies on Hadoop's role in healthcare systems and genomics data processing.Learn fundamental components such as MapReduce, HDFS, and YARNExplore MapReduce in depth, including steps for developing applications with itSet up and maintain a Hadoop cluster running HDFS and MapReduce on YARNLearn two data formats: Avro for data serialization and Parquet for nested dataUse data ingestion tools such as Flume (for streaming data) and Sqoop (for bulk data transfer)Understand how high-level data processing tools like Pig, Hive, Crunch, and Spark work with HadoopLearn the HBase distributed database and the ZooKeeper distributed configuration service.