

Problems with traditional systems What is Hadoop and what does it do? Hadoop Cluster Overview MapReduce Overview Hadoop Jobs and Tasks Introduction to Git and IntelliJ IDE
MapReduce detailed discussion Hadoop 1 vs 2 The Hadoop Distributed File System Running and debugging your first Big-Data program on Hadoop standalone mode
Installation and configuration of Hadoop in Pseudo-distributed mode Writing basic MapReduce programs in Java Accessing HDFS Programmatically
MapReduce job chaining Joining multiple data sources Performance optimization Interactive analysis with Impala
Hive: Batch processing SQL-like querying engine Pig: high level map-reduce programming language
Hbase: low latency random access analysis Sqoop: Data transfer tool between Relational database and HDFS
Spark: In memory data processing framework
NoSQL database Cluster management tools Discussion on Industry use cases of tools and techniques covered so far Road map for further studies