Transcript name: MapReduce – Part 2 – Submitting a job

English
The process of running a MapReduce job on Hadoop consists of 8 major steps. The first step is the MapReduce program you've written tells the JobClient to run a MapReduce job. This sends a message to the JobTracker which produces a unique ID for the job. The JobClient copies job resources, such as a jar file containing a Java code you have written to implement the map or the reduce task, to the shared file system, usually HDFS. Once the resources are in HDFS, the JobClient can tell the JobTracker to start the job. The JobTracker does its own initialization for the job. It calculates how to split the data so that it can send each "split" to a different mapper process to maximize throughput. It retrieves these "input splits" from the distributed file system. The TaskTrackers are continually sending heartbeat messages to the JobTracker. Now that the JobTracker has work for them, it will return a map task or a reduce task as a response to the heartbeat. The TaskTrackers need to obtain the code to execute, so they get it from the shared file system. Then they can launch a Java Virtual Machine with a child process running in it and this child process runs your map code or your reduce code. This lesson is continued in the next video.