6. Create a Data Pipeline with Apache Flume -> HDFS -> Spark

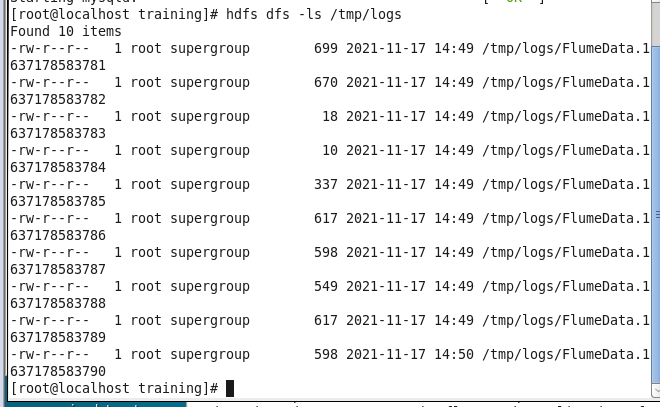
**FLUME to HDFS:**

hdfs dfs -mkdir /tmp/logs

vi /var/log/mysqld.log 🡪 log file that is getting transferred to HDFS

To transfer files to flume:

flume-ng agent --conf /etc/flume-ng/conf -f /etc/flume-ng/conf/flume.conf -Dflume.root.logger, console -n agent



Files have been transferred to HDFS.

**HDFS to SPARK:**

text = spark.read.text("hdfs://master:9000/flumedata.txt").show()

A picture containing background pattern

Description automatically generated

