

Evaluation of Erasure Coding & other features of Hadoop 3

Thursday 15 August 2019 14:05 (7 minutes)

Apache Hadoop is a set of 2 domains: data computation such as Spark, MapReduce, Flink, etc and data storage - HDFS. HDFS is a distributed file system. Current HDFS provides 3x replication for data redundancy and availability. But it has 200% storage overhead. However there is a big improvement in Hadoop 3 for replication which is Erasure Coding (EC).

Erasure Coding gives the same level of fault tolerance as 3x replication but with much less storage space.

My project aims to evaluate the performance of Erasure Coding.

Presenter: SEIDAN, Nazerke