

Evaluation of Erasure Coding & other features of Hadoop 3

Nazerke Seidan

Supervisors: Emil Kleszcz, Zbigniew Baranowski

Project Motivation

**How to reduce storage overhead
without affecting the performance
of Big Data Processing?**



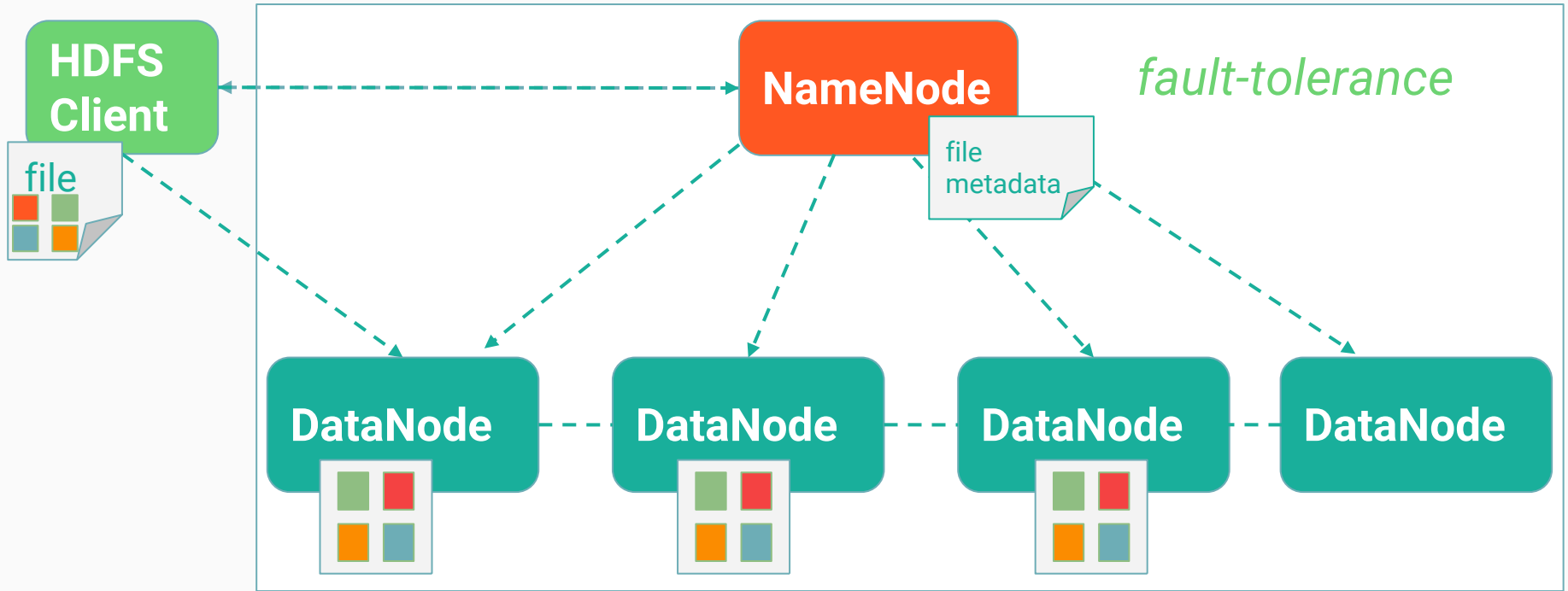
Hadoop

Compute



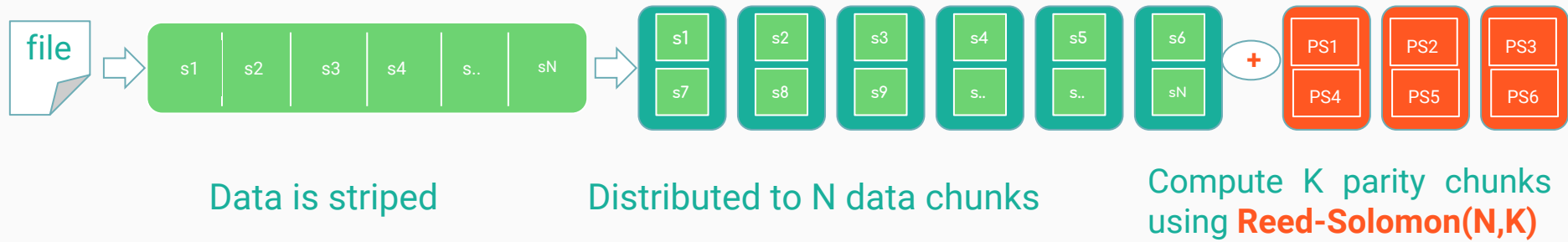
Storage
(HDFS)

Current HDFS Replication



3x replication: 200% storage cost

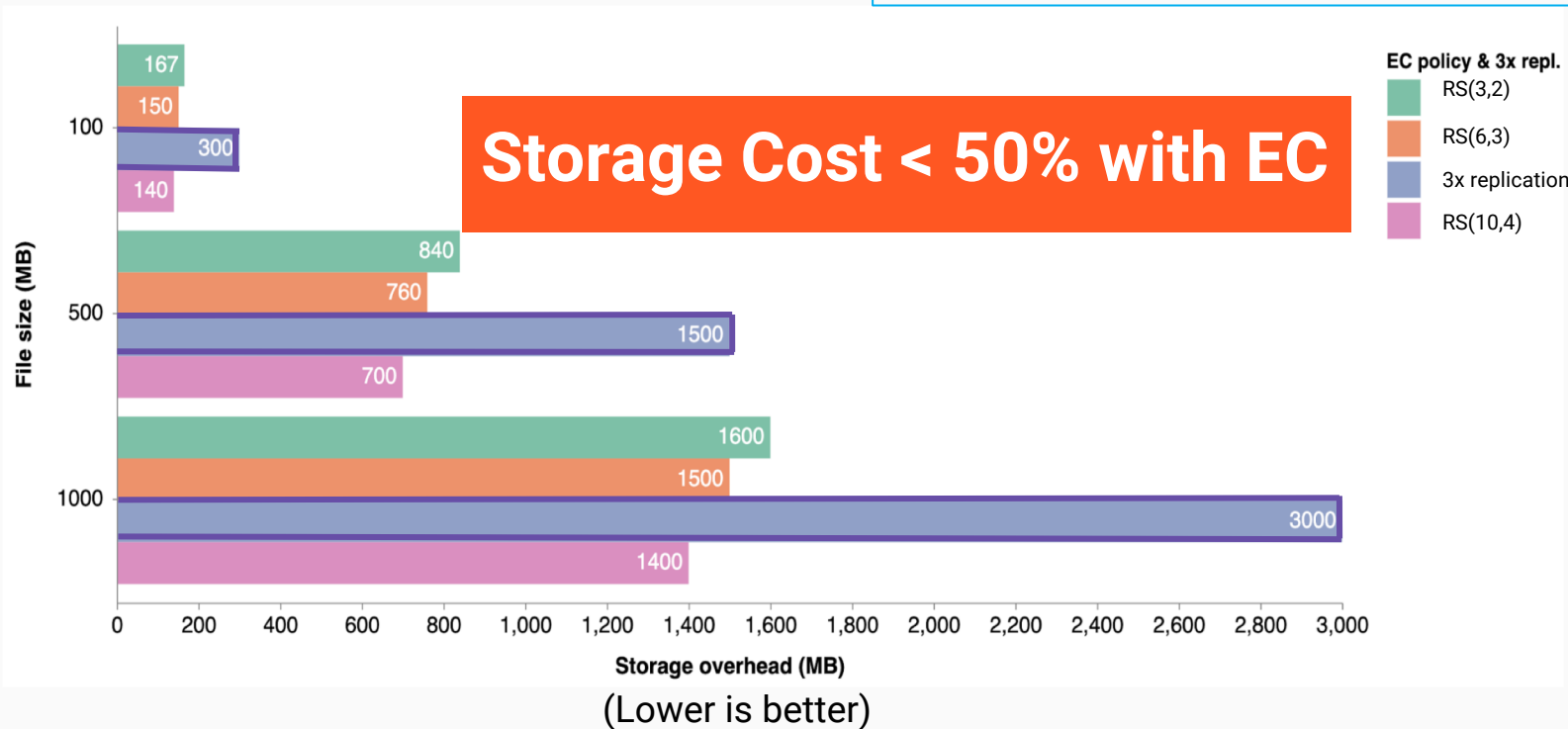
Solution: Erasure Coding (EC)



- * Similar to RAID 5/6 concept
- * Uses Reed-Solomon (RS) algorithm
- * RS (N,K) where N - data chunks, K - parity chunks
- * Generates K parity chunks based on N data chunks
- * Tolerates up to K failures

Evaluation Result

Our production-like cluster:
16 machines, 48 drives/machine
5.5TB/drive

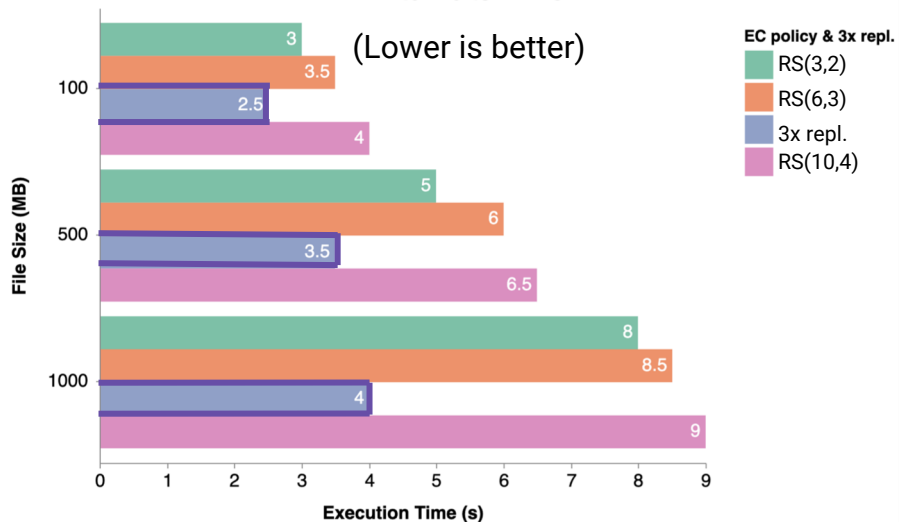


Evaluation Result

Intel® ISA-L

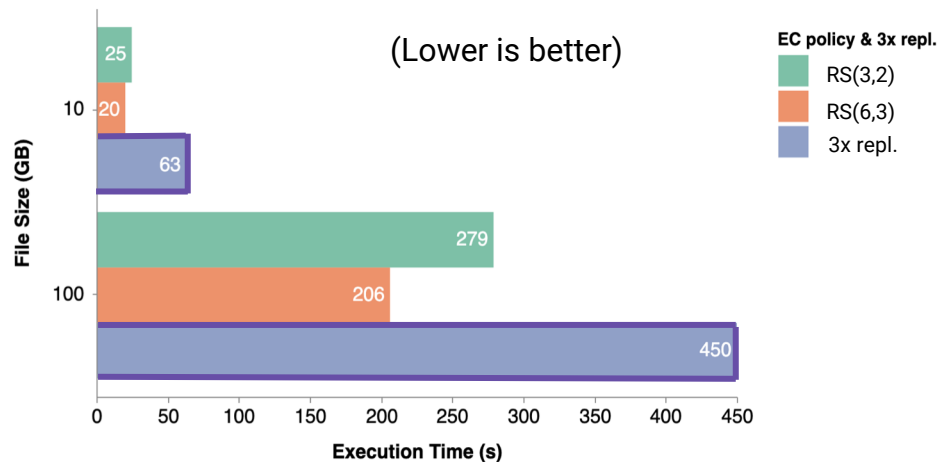
Write file to HDFS

(Lower is better)



Read file from HDFS

(Lower is better)

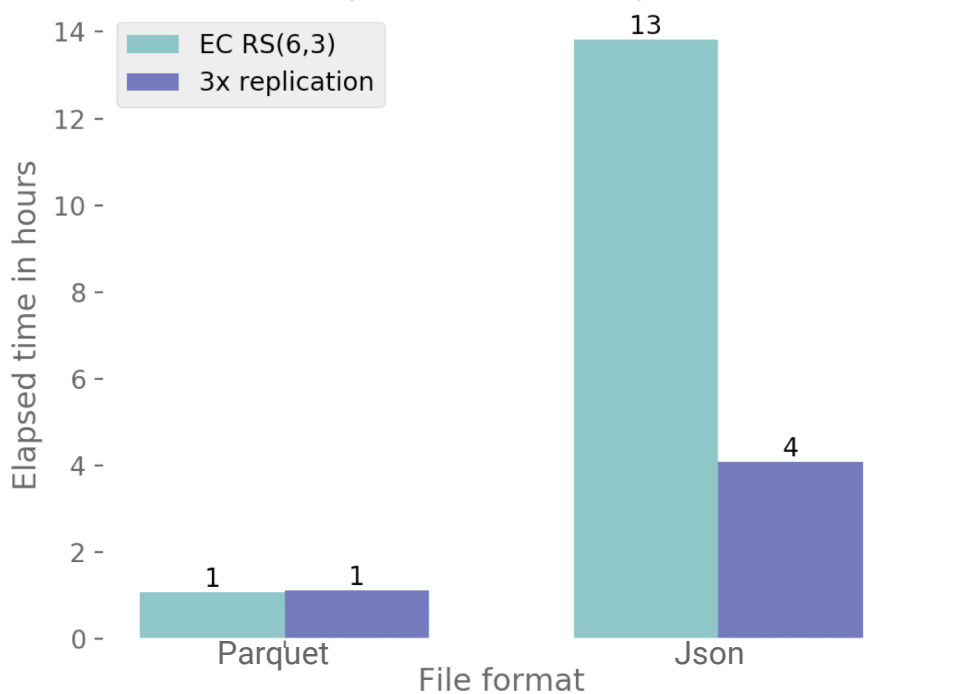


40-50% slower with EC

2x faster with EC

Evaluation Result

TPC-DS benchmark test on 3X & EC RS(6,3), 1TB
(Lower is better)



* TPC-DS: decision support benchmark



**Unoptimized file formats
should be avoided with EC**

**Performance with
optimized file formats is
the same for both
replication policies**

What Next?

Project Impact:

- * Erasure Coding will be applied to CERN Hadoop Cluster in production

Evaluate other features of Hadoop 3:

- * Triple NameNode High Availability (HA) (ongoing)
- * HDFS Router-based Federation

Contact:

Nazerke Seidan



github.com/nazerkebs



linkedin.com/in/nazerkes



seinaz1997@gmail.com

Acknowledgements

To my supervisors Emil Kleszcz,
Zbigniew Baranowski and CERN IT-DB team

To Rakesh Radhakrishnan for better understanding
of the EC. Sr.Software Engineer,
Hadoop PMC Member, Intel Corp.
<http://people.apache.org/~rakeshr/>

To CERN openlab organizers