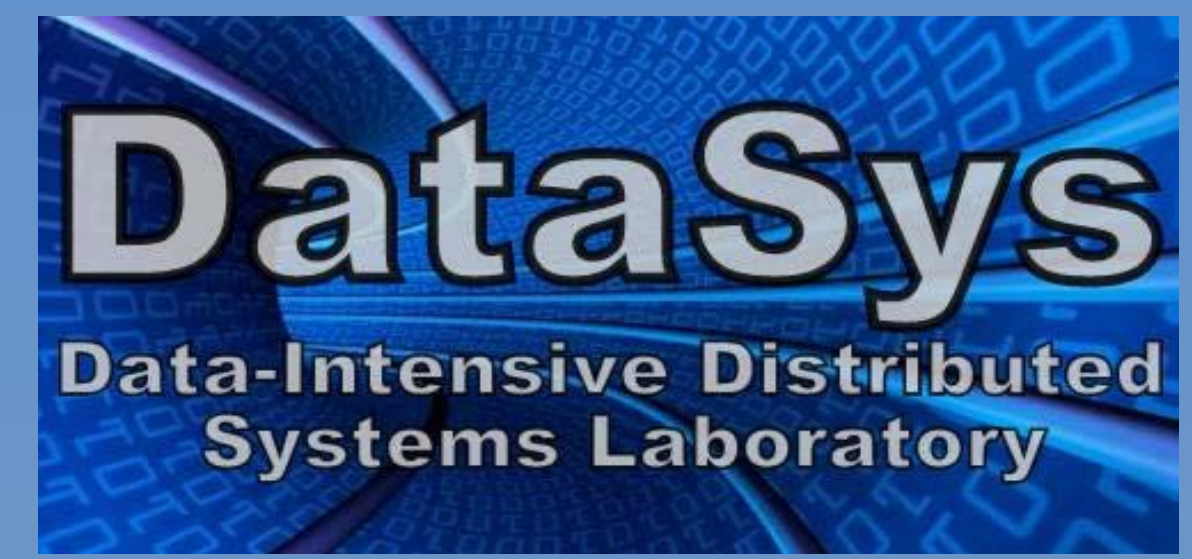


FusionFS: a distributed file system for large scale data-intensive computing



Dongfang Zhao¹, Chen Shou¹, Zhao Zhang²
 Iman Sadooghi¹, Xiaobing Zhou¹, Tonglin Li¹, Ioan Raicu^{1,3}
¹Department of Computer Science, Illinois Institute of Technology
²Department of Computer Science, University of Chicago
³Mathematics and Computer Science Division, Argonne National Laboratory

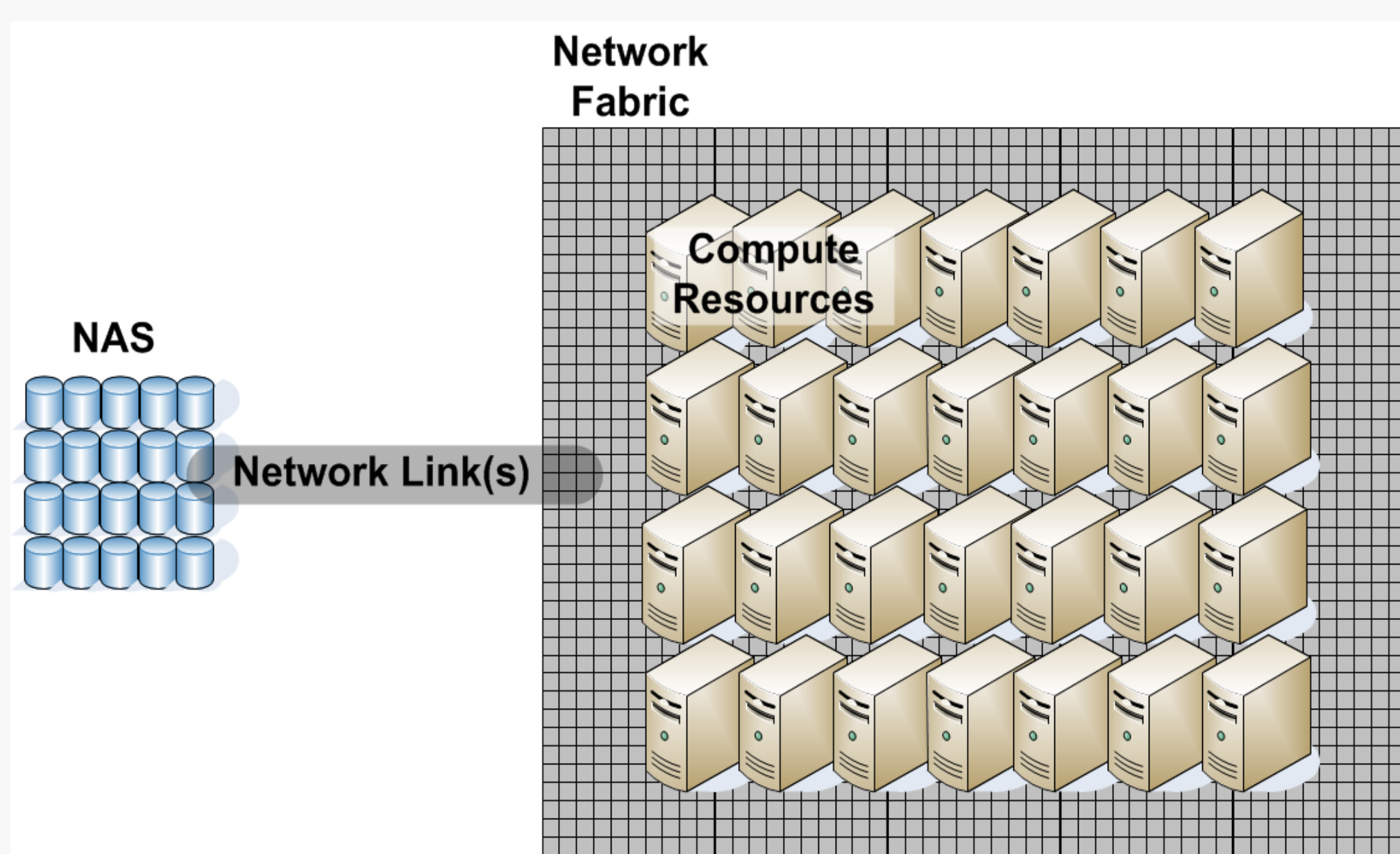


Goal

Develop both theoretical and practical aspects of building distributed file systems scalable to exascale supporting millions of nodes and billions of concurrent IO requests

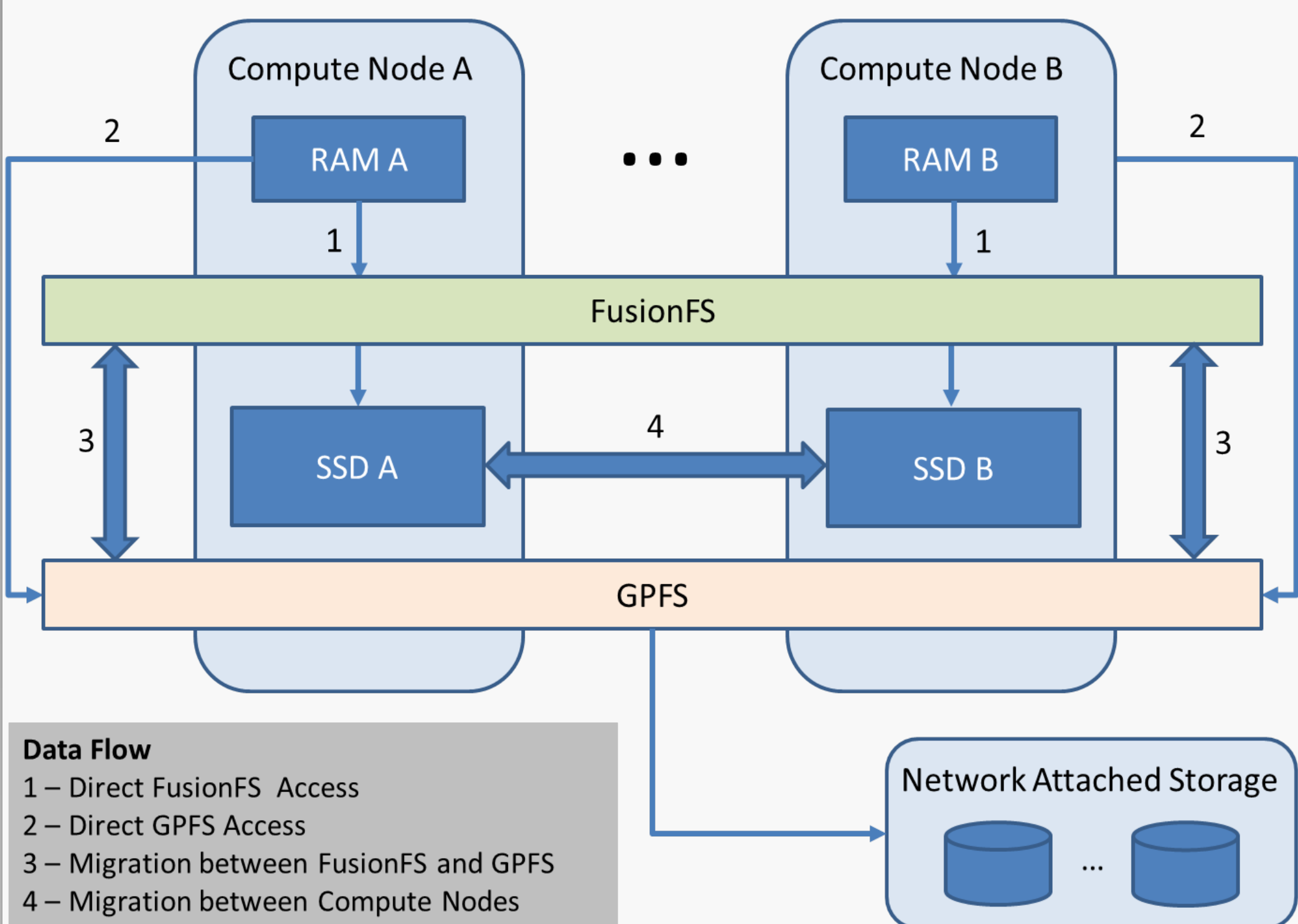
Motivation

Current architecture (i.e. compute nodes are remotely connected to storage nodes) would unlikely scale well at exascale

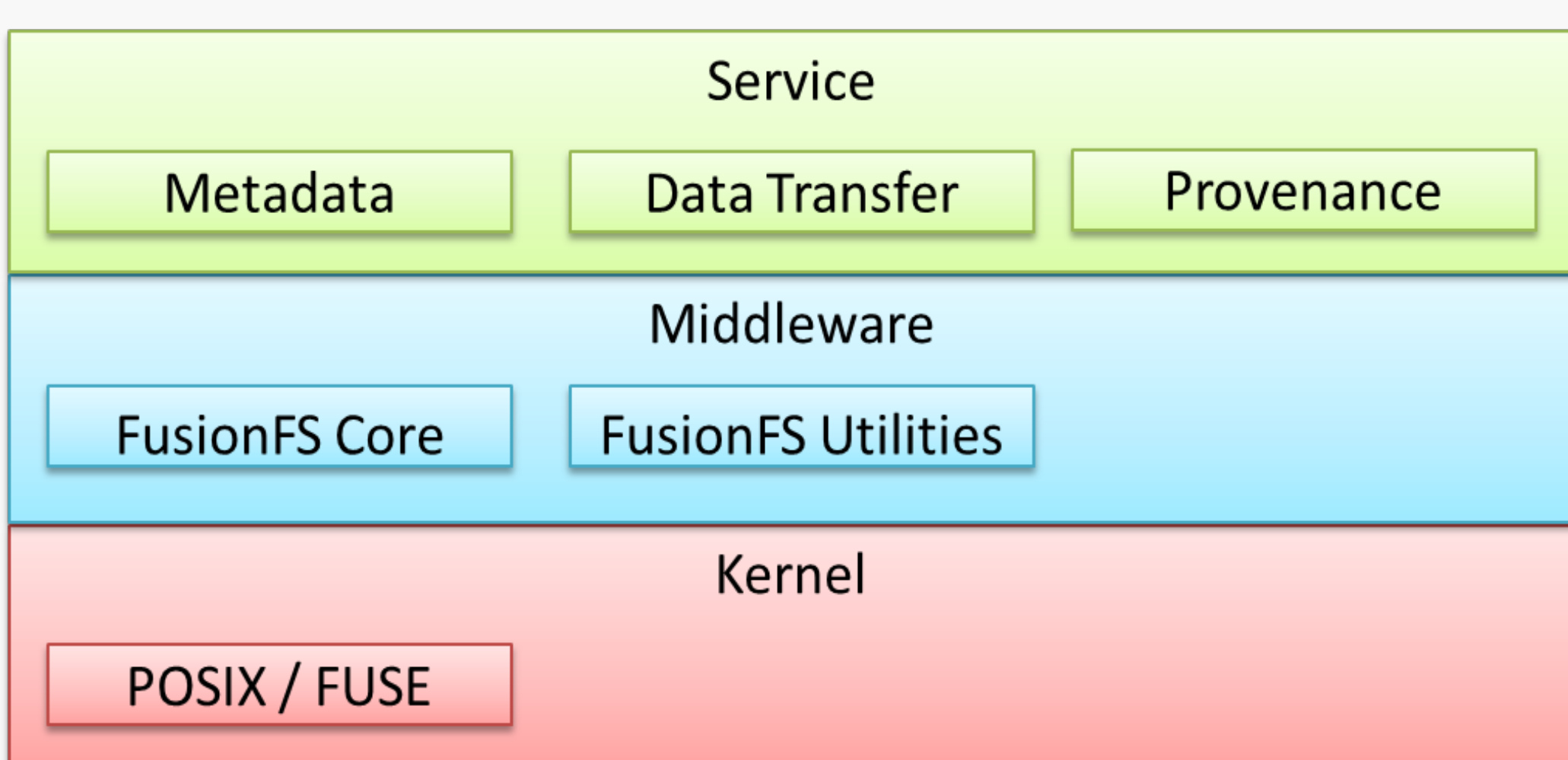


Proposed Architecture

❖ Distributed file system coexists on compute nodes



Software Stack



References

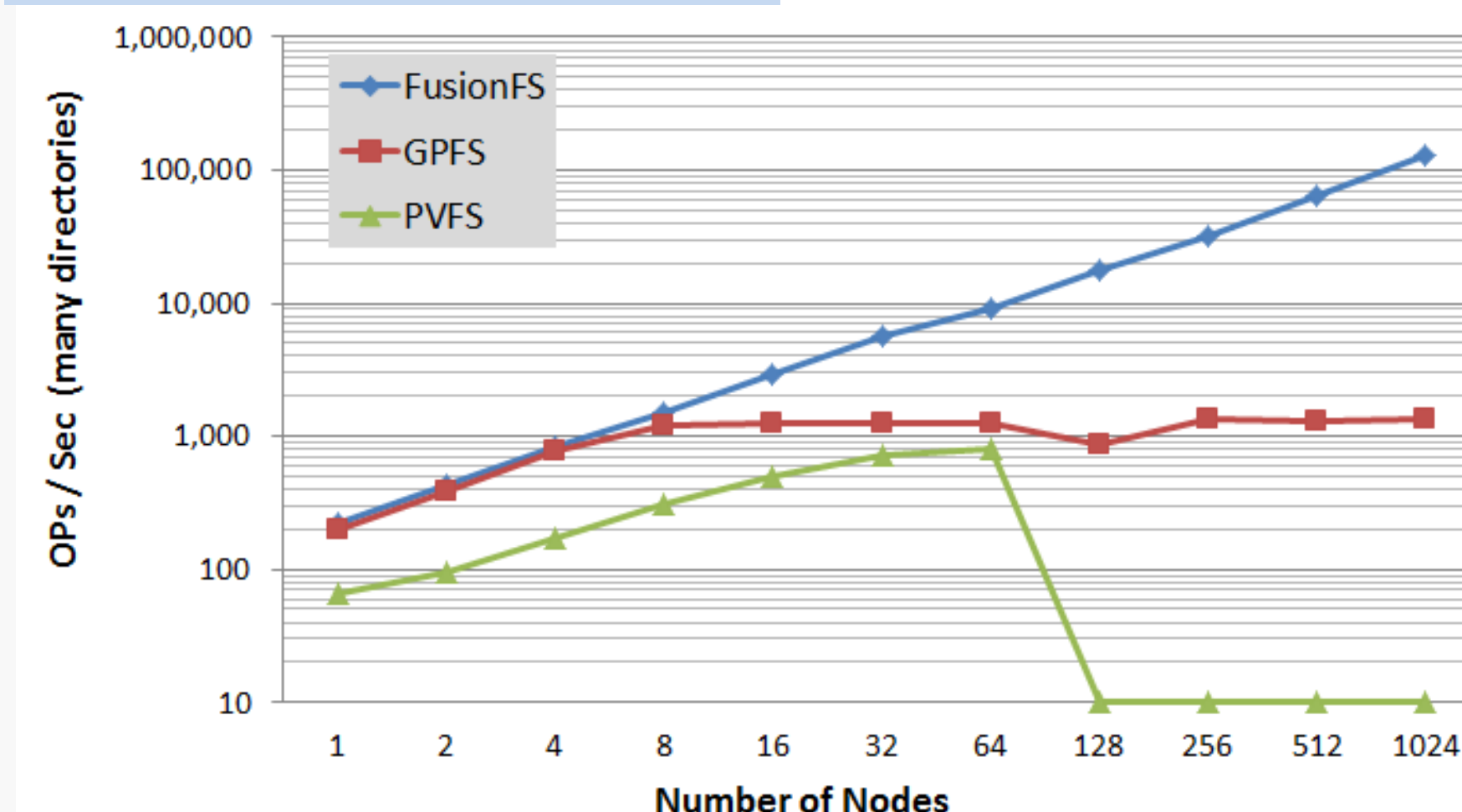
- [1] FusionFS project website:
<http://datasys.cs.iit.edu/projects/FusionFS/index.html>
 [2] Ioan Raicu, Ian Foster and Pete Beckman. Making a Case for Distributed File Systems at Exascale, *ACM Workshop on Large-scale System and Application Performance (LSAP)*, 2011

Acknowledgement

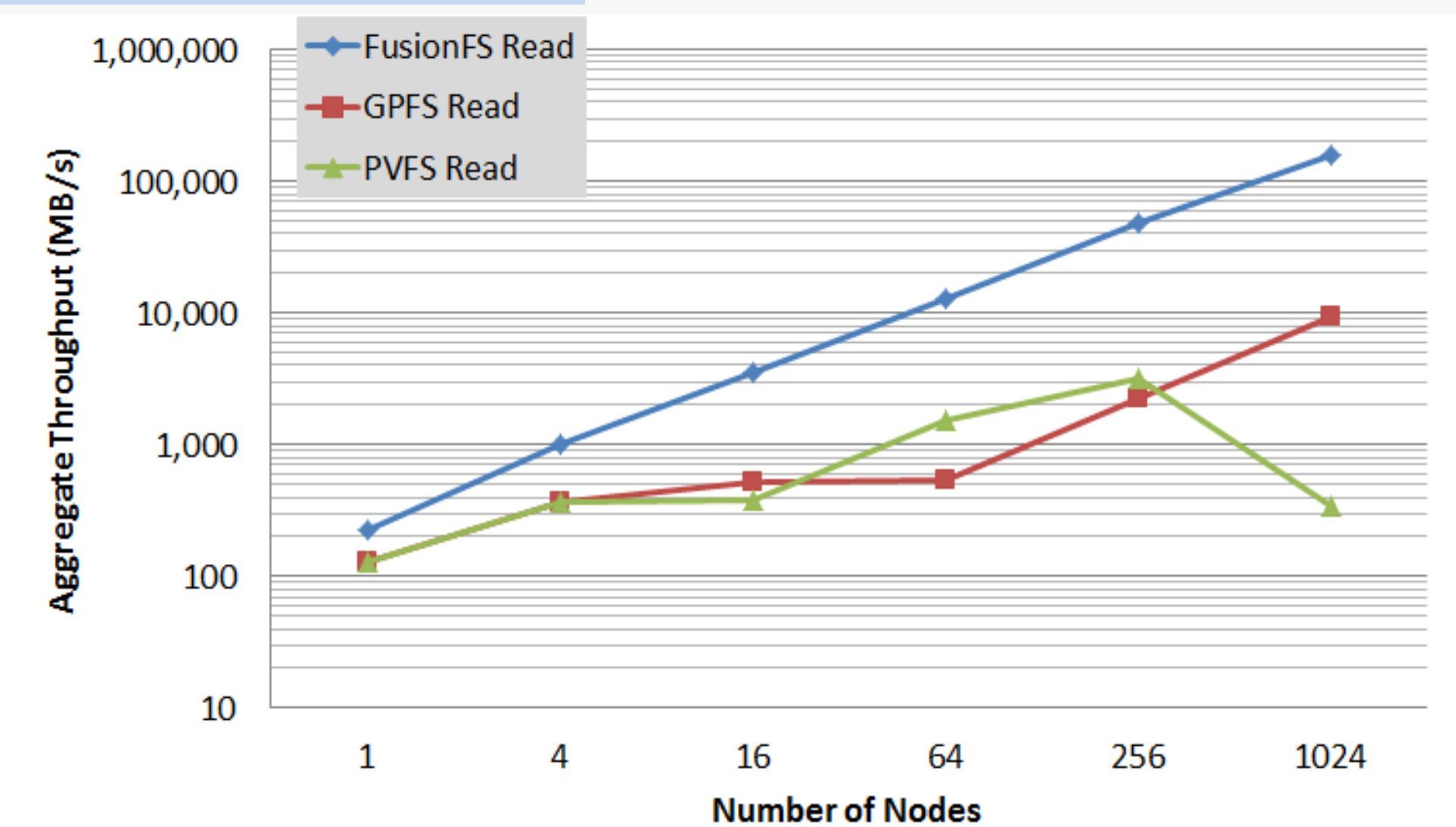
This work is supported by NSF grant OCI-1054974

Performance on Blue Gene/P

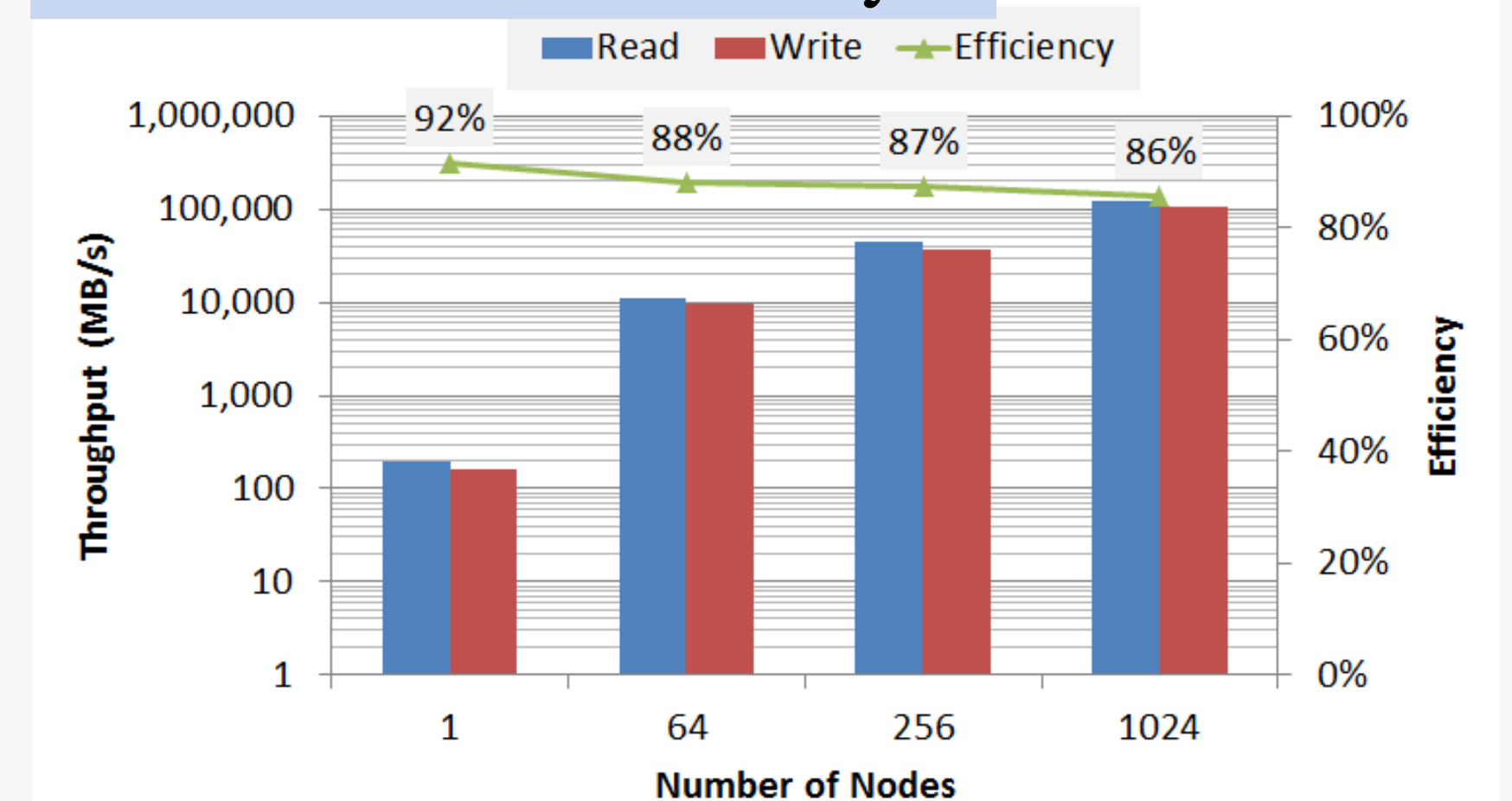
Metadata Ops/sec



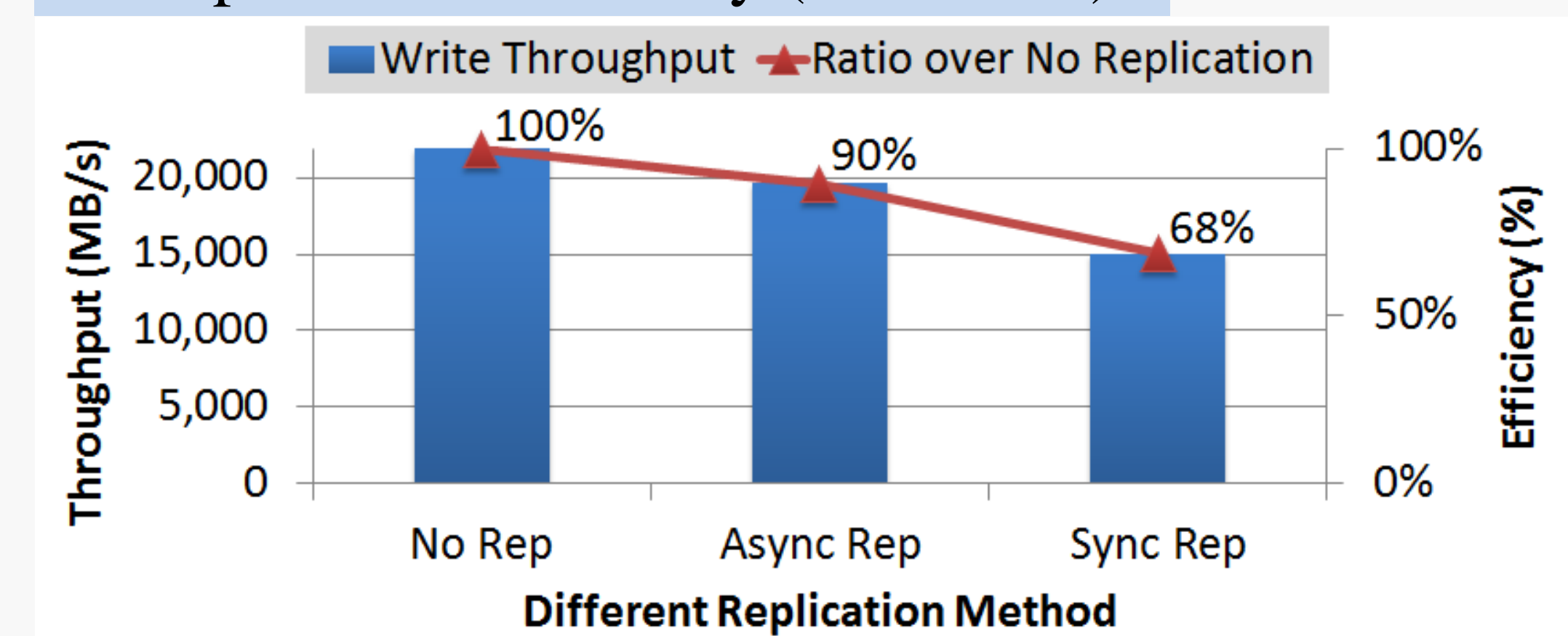
Data Throughput



Provenance Efficiency

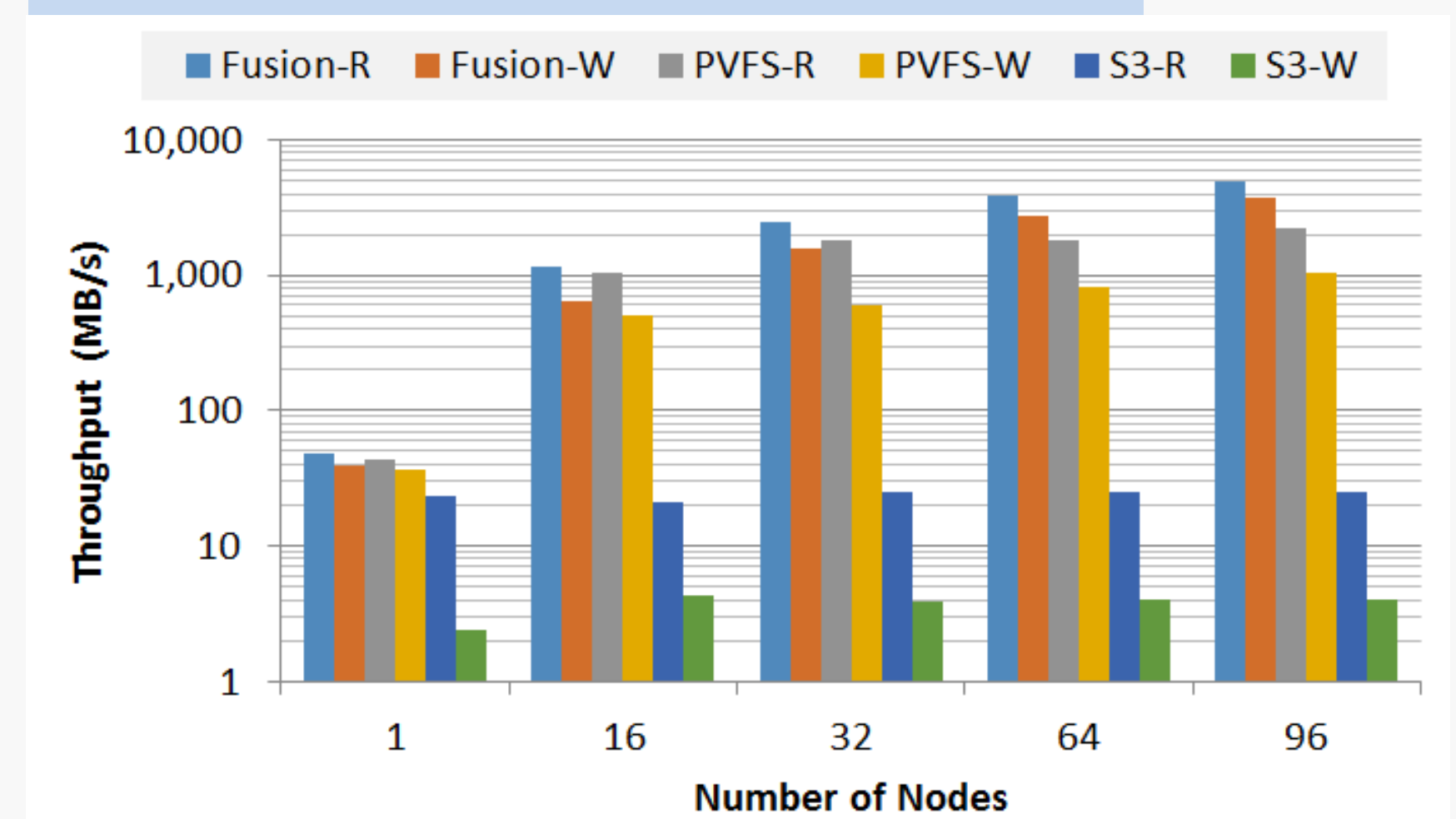


Replication Efficiency (128 nodes)

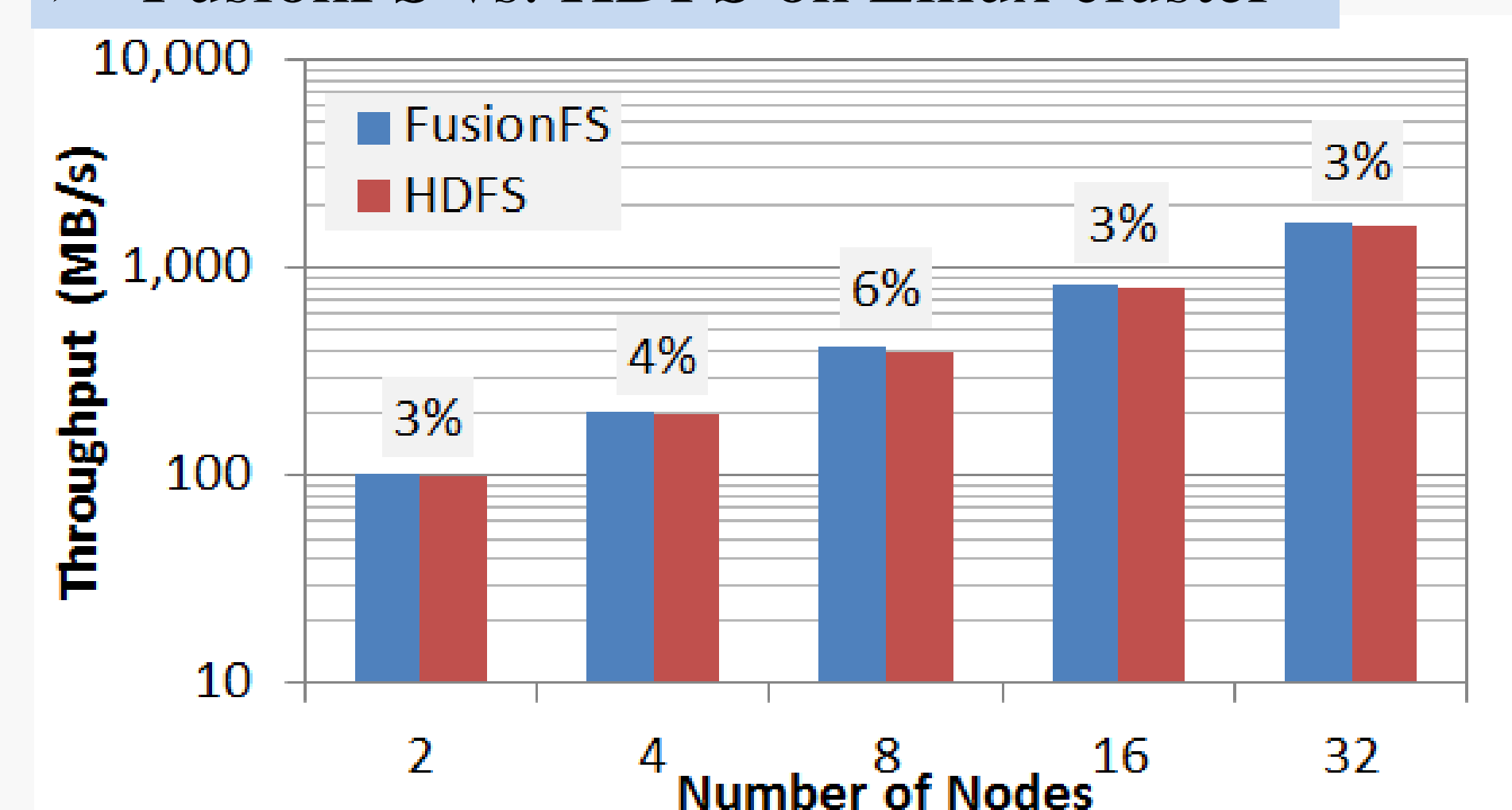


Performance on Amazon cloud and Linux cluster

Amazon m1.medium instance

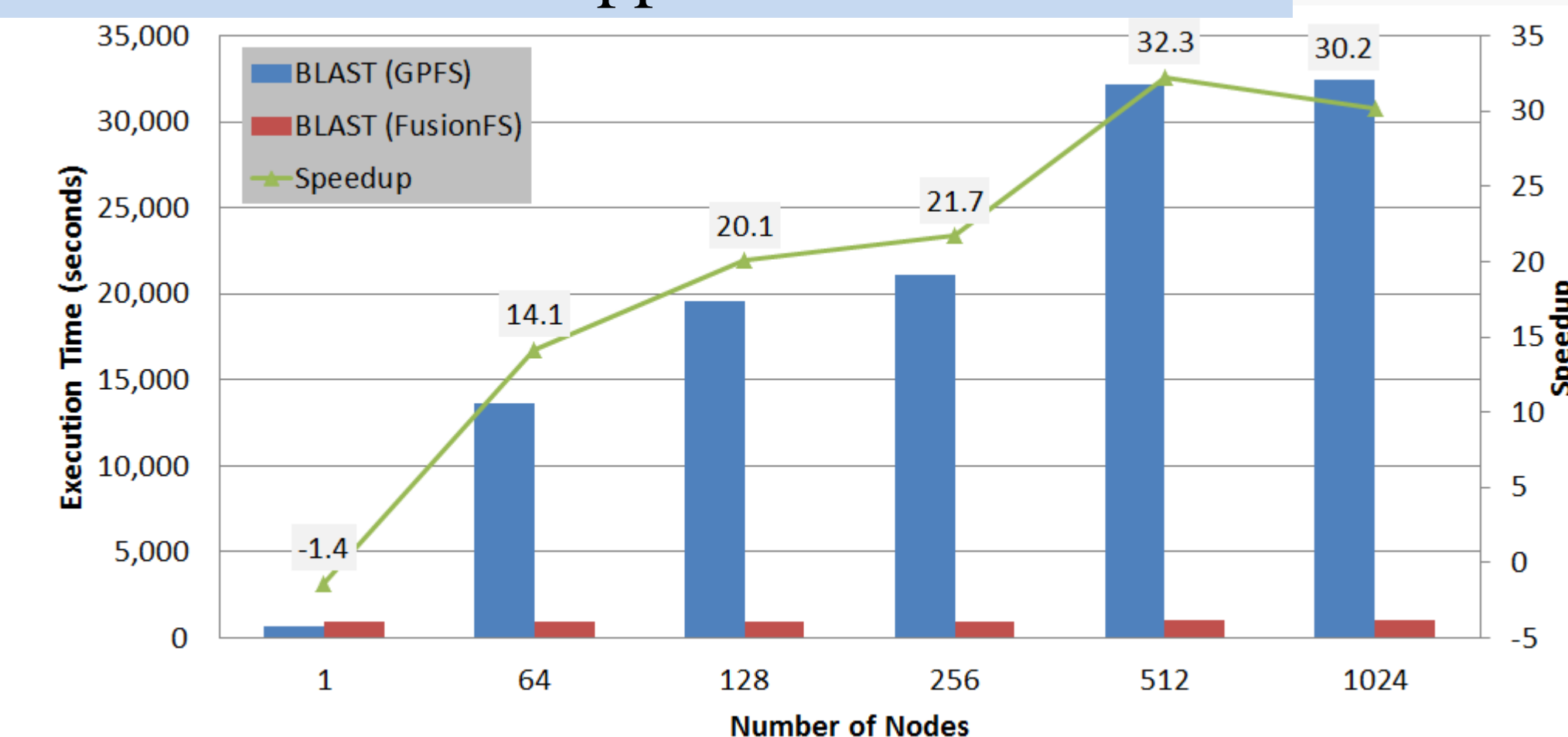


FusionFS vs. HDFS on Linux cluster

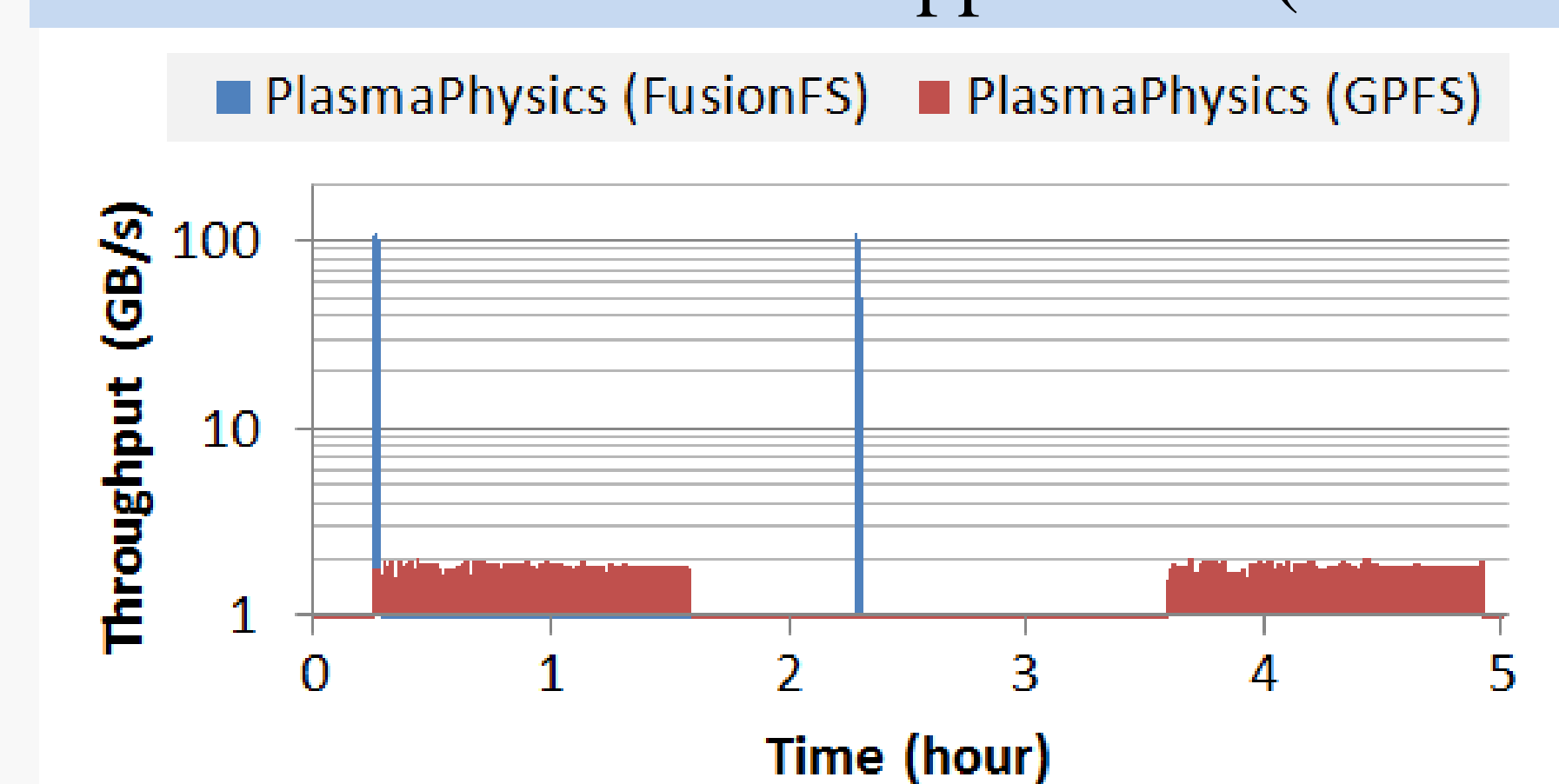


Performance on Applications

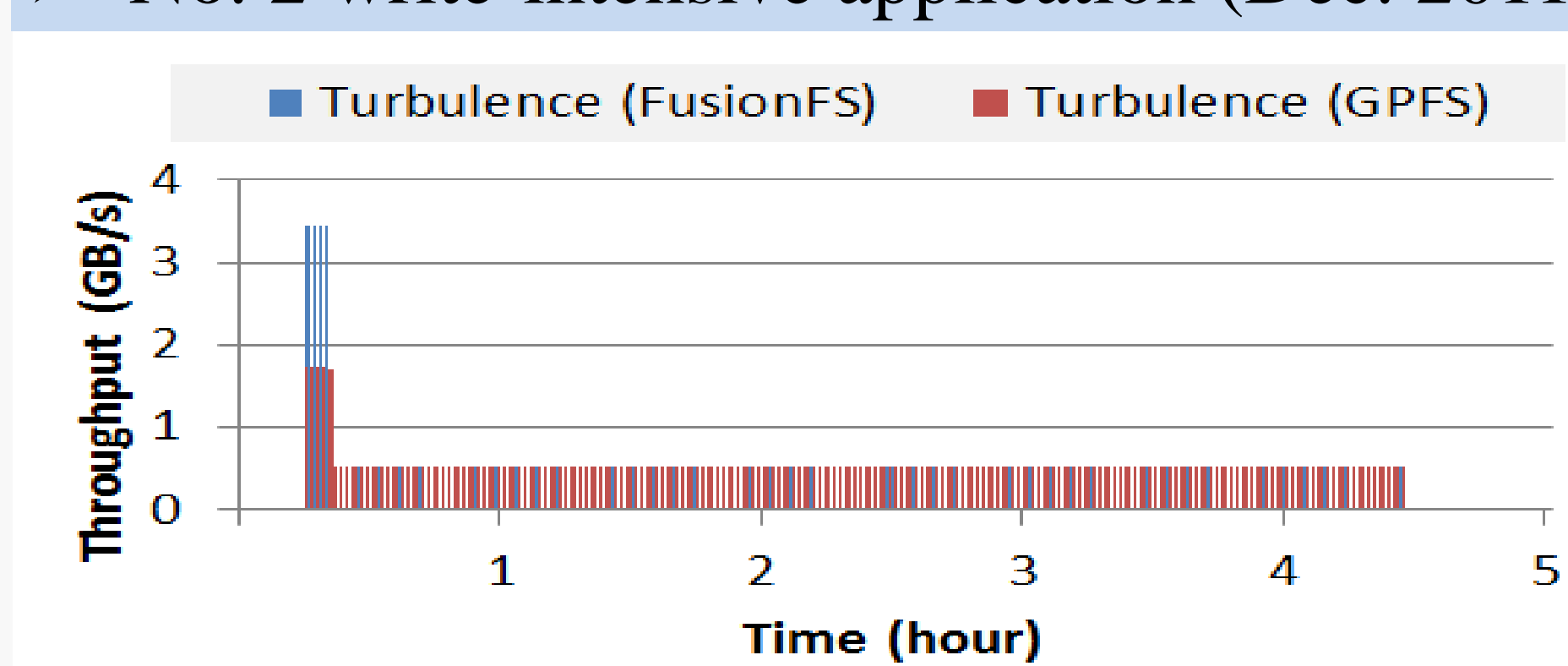
Bioinformatics application: BLAST



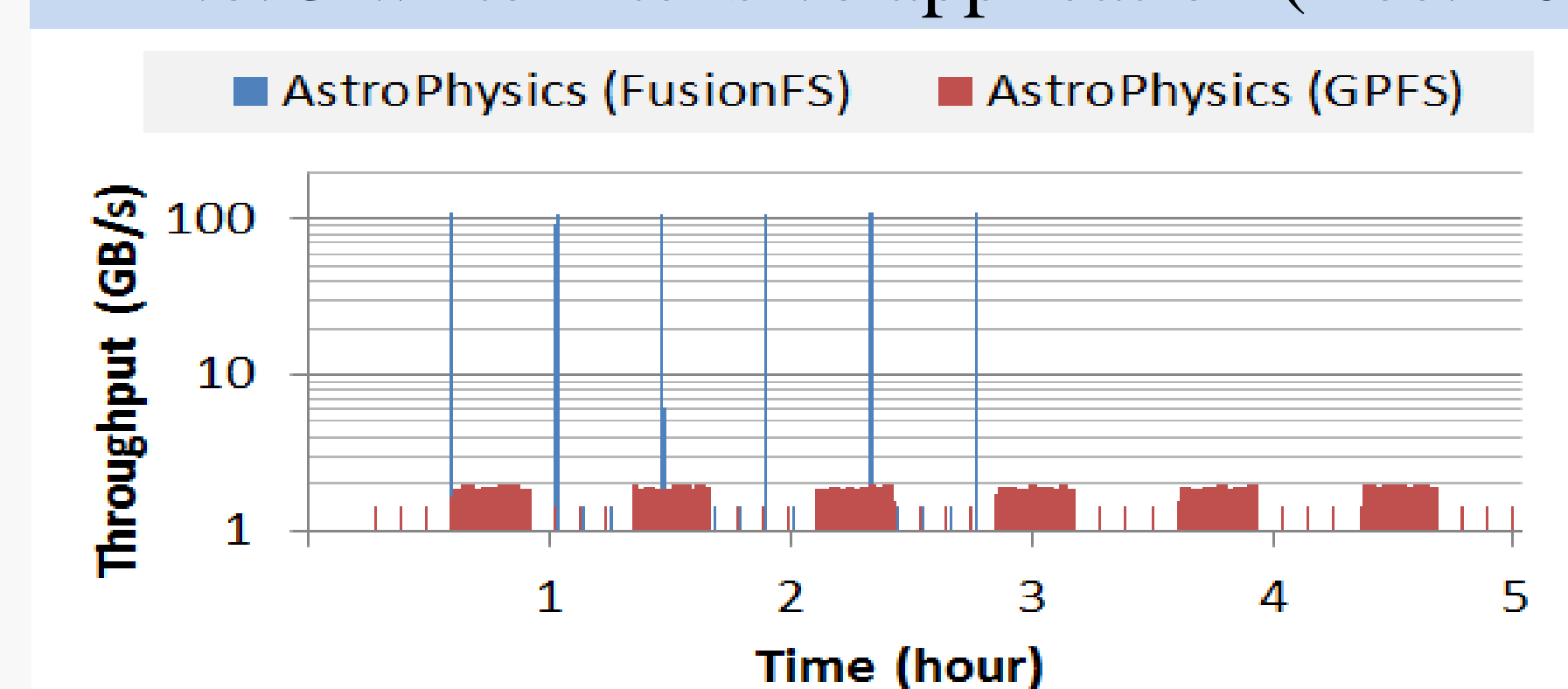
No. 1 write-intensive application (Dec. 2011)



No. 2 write-intensive application (Dec. 2011)



No. 3 write-intensive application (Dec. 2011)



Conclusion and Future Work

- FusionFS delivers significantly higher throughput of data and metadata than other leading file systems
- FusionFS provides a completely distributed and loosely-coupled metadata management
- FusionFS maximizes the data locality for file writes, making it a perfect fit for checkpoints
- FusionFS supports different replication semantics, as well as efficient data provenance
- We are scaling FusionFS to O(10K) nodes, and developing a library to bypass FUSE