Accelerating Big Data Processing in the Cloud with Scalable Communication and I/O Schemes

Shashank Gugnani, Dhabaleswar K. Panda (Advisor), The Ohio State University

Overview

Current Trends in Big Data

- Huge increase in cloud deployments running Big Data analytics
- Analytics performed on data stored in cloud storage
- System and job sizes constantly increasing
- High-performance solutions for Big Data in the cloud essential

Importance of Big Data in Cloud

- Inherent flexibility and scalability
- Tremendous cost saving
- Built-in reliability and fault-tolerance

Network Communication Bottlenecks

- Not aware of topology and locality
- Slow TCP-based

Scalability Issues

- Cloud storage solutions have limited scalability
- Limited number of gateway or proxy servers limits operation throughput

Consistency Issues

- Cloud storage systems typically provide Eventual Consistency (EC)
- EC is not sufficient for traditional applications expecting POSIX-like consistency

Proposed Designs

High-performance communication[1]

- Use of RDMA-based low latency communication
- Use of SR-IOV hardware virtualization with VMs

Topology-aware communication[2]

- MapReduce-based automatic topology detection
- Locality and topology-aware communication and scheduling

High-performance Cloud Storage[3]

- RDMA-based communication
- Re-designed scalable architecture with client-based replication

POSIX-like consistent Cloud Storage

- Proposed use of atomic operations to provide consistency
- Implemented 2PC for write operations

Contributions

- Near-native performance (< 9% overhead) for applications in virtualized environments
- Scalable automatic topology detection
- Efficient topology and locality-aware communication
- High-performance and consistent cloud storage
- Ability to run version control, database, and big data applications directly on cloud storage

Publications

[1] Performance Characterization of Hadoop Workloads on SR-IOV-enabled Virtualized InfiniBand Clusters. (Gugnani et al, BDCAT '16)

[2] Designing Virtualization-aware and Automatic Topology Detection Schemes for Accelerating Hadoop on SR-IOV-enabled Clouds. (Gugnani et al, CloudCom '16)

[3] Swift-X: Accelerating OpenStack Swift with RDMA for Building an Efficient HPC Cloud. (Gugnani et al, CCGrid '17)

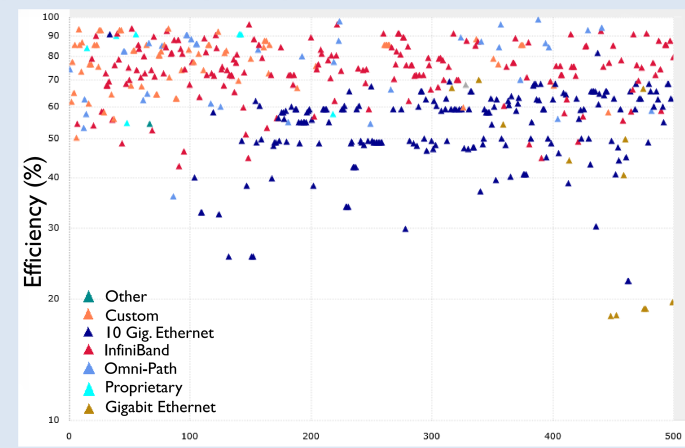
More Information

- http://hibd.cse.ohio-state.edu/
- Proposed designs will be released soon!

Challenges

Slow Network Communication

- TCP-based communication causes bottlenecks
- Each message transfer leads to context switches
- Software-based network virtualization leads to further slowdown



Efficiency data from Top500 Supercomputers

Inefficient Communication

- For large-sized clusters, topology-aware communication is paramount
- Existing topology-aware designs in Hadoop are not optimized for cloud environments
- No service that can automatically detect cluster topology and expose it to Hadoop

Process Location		Number of Hops	Latency (us)
Intra-Rack	Inter-Chassis	0 Hops in Leaf Switch	1.57
	Intra-Chassis	1 Hop in Leaf Switch	2.04
Inter-Rack	-	3 Hops in Leaf Switch	2.45
		5 Hops in Leaf Switch	2.85

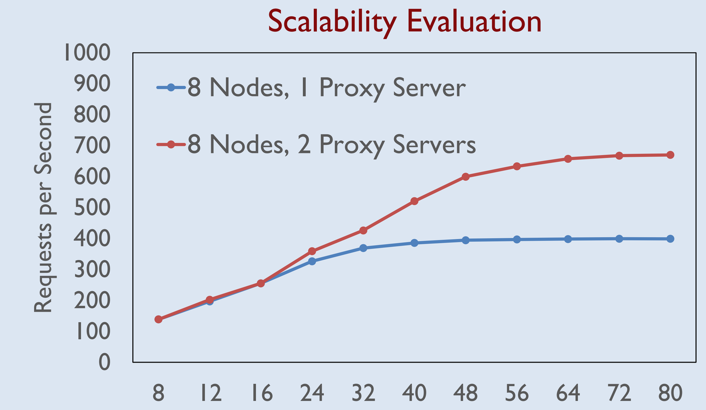
Reference: https://confluence.pegasus.isi.edu/download/attachments/5242944/topology-aware-poster.pdf

Communication Data from TACC Ranger System

Limited Scalability in Cloud Storage

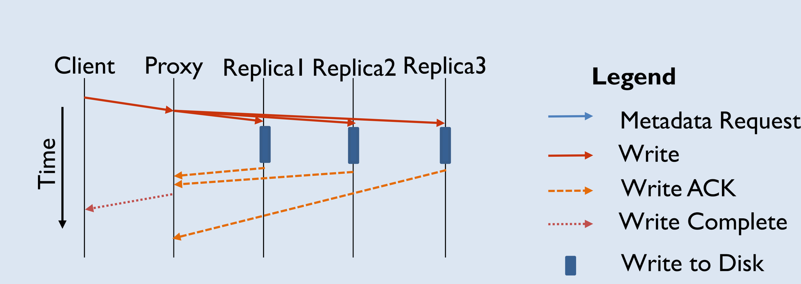
- Proxy server design in Swift limits throughput since all operations are routed through the proxy server
- Server-side replication limits scalability
- Network communication is slow TCP-based

Scalability Evaluation



Consistency Issues

- Traditional applications reliant on POSIX-like consistency
- Cloud storage solutions provide Eventual Consistency (EC)
- Application migration to the cloud is not straightforward
- Consistency guarantees are required



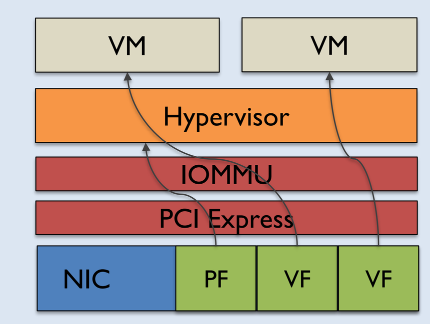
Default Write Design

Consistency not guaranteed

Proposed Designs

Modern Networking Protocols

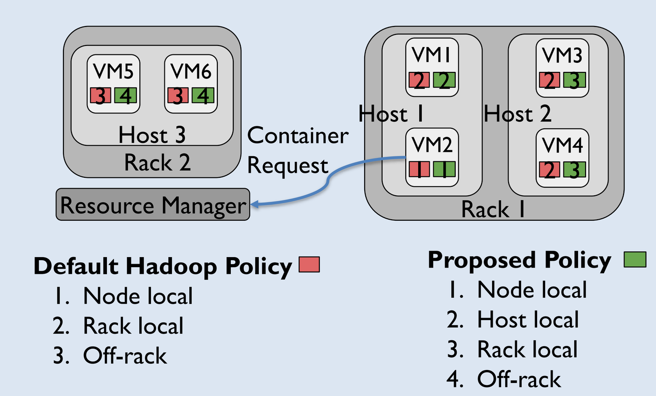
- InfiniBand and RoCE provide RDMA-based efficient communication
- SR-IOV offers hardware-based network virtualization
- With SR-IOV, VMs can directly access the network adapter
- Comprehensive evaluation of Hadoop workloads on IB clusters with SR-IOV



SR-IOV Overview

Topology-aware Communication

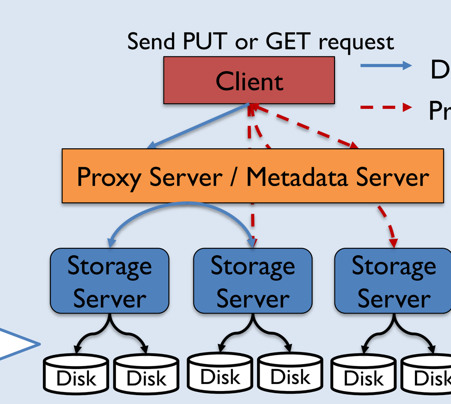
- Automatic topology detection module can detect topology changes during runtime
- Maximize communication between co-located VMs
- Allocate Containers and Map tasks on a co-located VM before other VMs



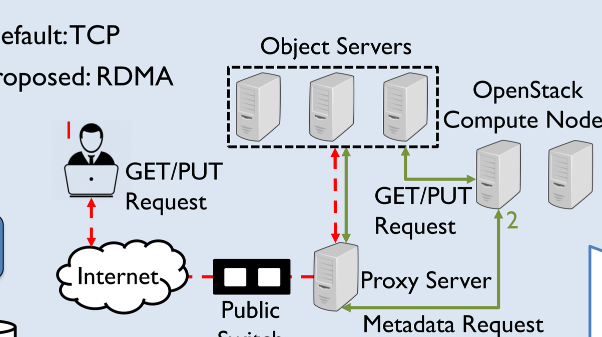
Proposed Container Allocation Policy

Scalable Cloud Storage: Swift-X

- Use proxy server only as a metadata server
- Client-based replication for scalability
- RDMA-based communication for high-performance
- Non-blocking semantics for efficient overlap between communication and I/O



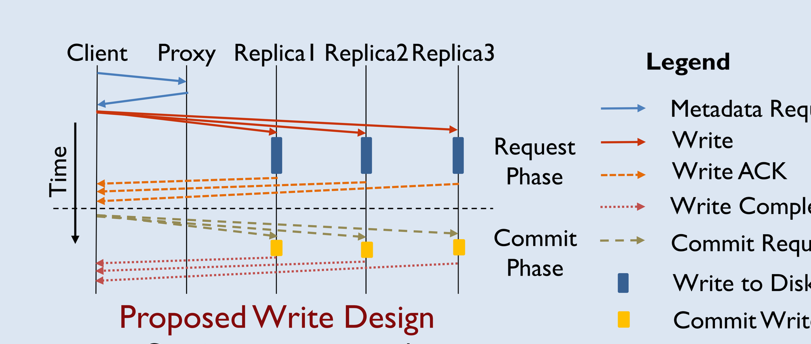
Swift Design Overview



Proposed Usage Scenarios

POSIX-like consistent Cloud Storage

- Atomicity as a way to guarantee consistency
- Two-phase commit (2PC) for atomic write operations
- Client-side caching to improve read/write performance
- Compatibility with HDFS API: MapReduce workloads can directly run on cloud storage



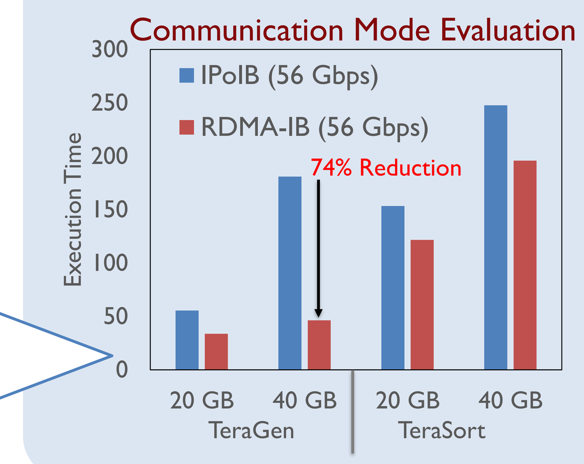
Proposed Write Design

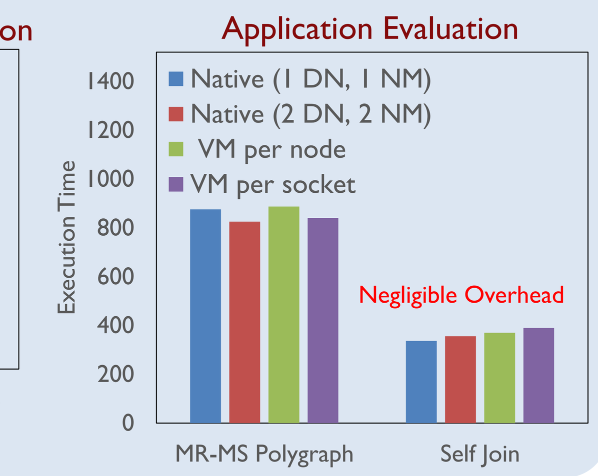
Consistency guaranteed

Results

Evaluation on Virtual Cluster

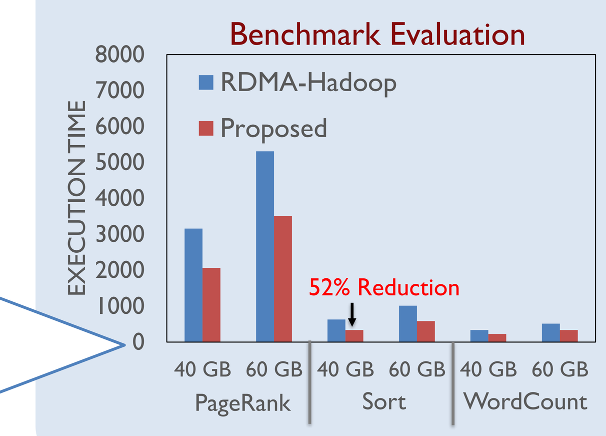
- Less than 9% overhead for applications compared to native execution
- Selecting correct VM subscription policy can deliver near-native performance
- Up to 74% improvement for TeraGen, 21% for TeraSort for RDMA over IPoIB

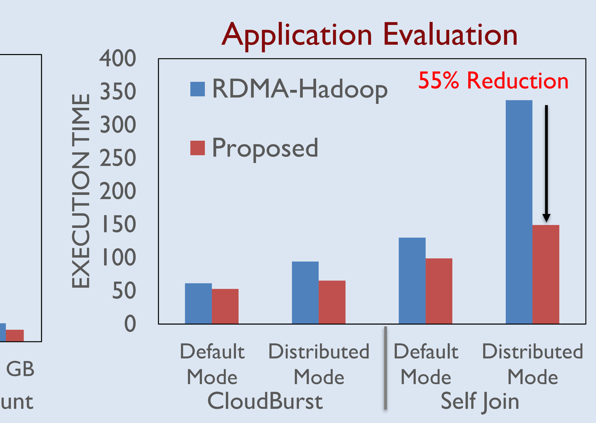




Evaluation with RDMA-Hadoop

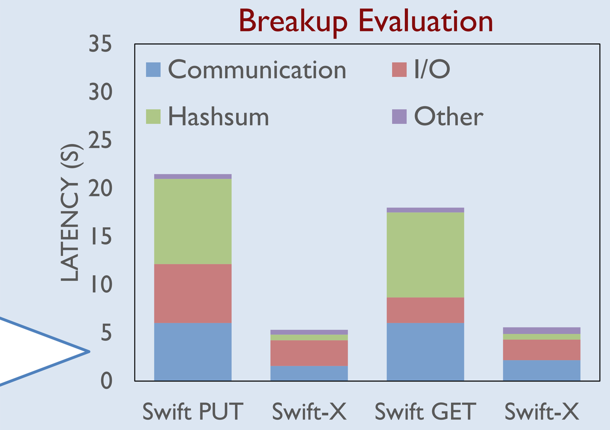
- Up to 52% improvement over RDMA-Hadoop for benchmarks
- Up to 55% improvement over RDMA-Hadoop for applications
- Proposed design delivers the best performance and fault-tolerance

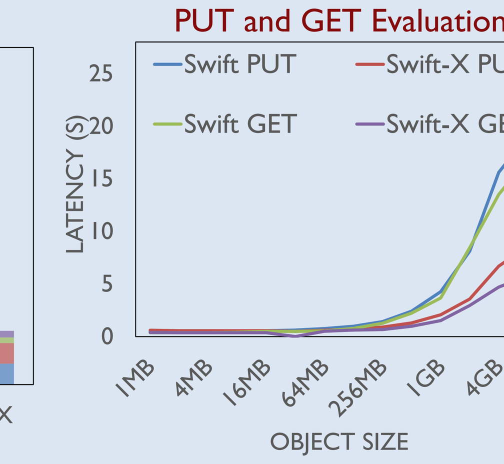




Evaluation with OpenStack Swift

- Up to 47% and 66% reduction in PUT and GET latencies
- Communication time reduced by up to 3.8x for PUT and up to 2.8x for GET
- Up to 7.3x improvement in read throughput for cloud storage





Evaluation with SwiftFS and HDFS

- Up to 83% improvement over SwiftFS
- Up to 64% improvement over HDFS
- With HDFS, data is copied from Swift
- Best performance and guaranteed consistency