I/O-Efficient MapReduce in the Cloud

Michael Conley, Amin Vahdat, and George Porter

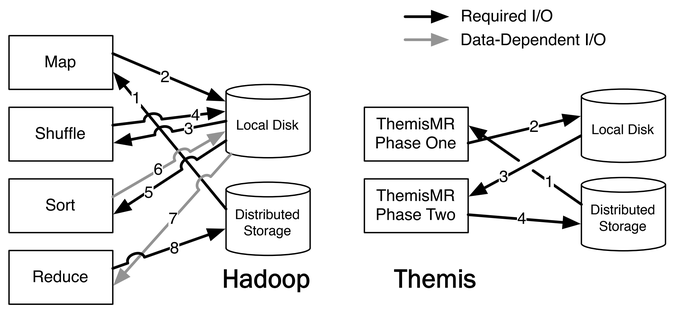
{mconley,vahdat,gmporter}@cs.ucsd.edu

Themis: I/O Efficient MapReduce

Goal: use the principles and ideas developed in TritonSort (NSDI 2011) to build a fast, general-purpose data processing system for data-intensive applications

- Specifically targeting workloads larger than memory
  - Outside the scope of modern, highly-efficient frameworks, e.g. Spark
  - I/O operations unavoidable => optimize I/O

Key Idea: The "2-IO" Property



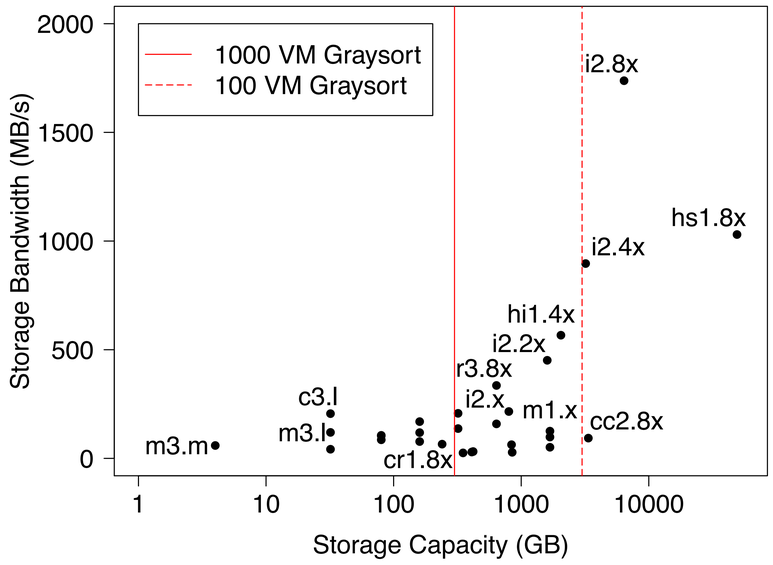
- Systems that read and write each record exactly twice (theoretical minimum) display **2-IO property**; fewer I/Os ⇒ higher performance
- Eliminate task-level fault tolerance: for moderately-sized clusters, I/O reduction means restart can provide higher throughput
- Central management of network and storage I/O

Moving to the Cloud

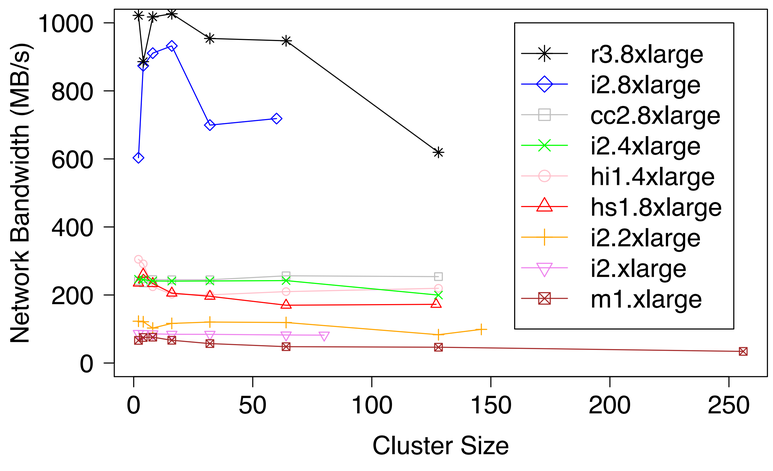
- **Goal:** Maintain efficiency in shared environment
- TritonSort and Themis designed for private cluster
  - Full control of dedicated hardware
- Realistically an application will be co-located
  - Public cloud for economies of scale
  - Private data center to maximize resource utilization

Case Study: Amazon EC2

- Amazon Elastic Compute Cloud (EC2) provides on-demand access to hundreds of virtual machines
  - 37 VM flavors to choose from
  - Wildly different performance characteristics

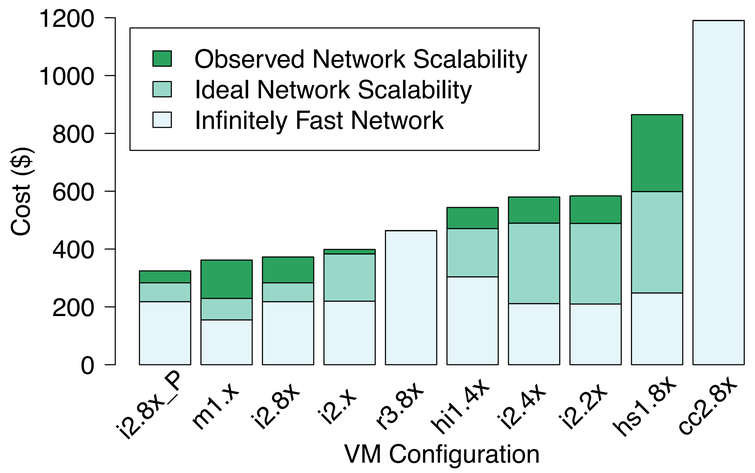


- Further, lack of isolation in the network affects performance at scale
  - Benchmarking a single VM is not enough



Sorting 100 TB on EC2

- **Question:** Which instance type should you choose?
  - Current best practice – use many small, cheap VMs
- **Result:** using fewer large, expensive VMs is better!



- More expensive VMs have "Enhanced Networking"
  - Can specify a launch placement group
  - Amazon will attempt to place VMs to ensure high bisection bandwidth
  - Trades scale for network isolation
  - i2.8xlarge with placement groups is the cheapest
    - But per-VM it is the most expensive ($6.82/hr)
    - Compare with m1.xlarge ($0.35/hr)

Themis on EC2

- Themis sorts 100 TB on 178 i2.8xlarge VMs in **888 s**
  - More than **4x faster** than current world record
  - Using **11x fewer** servers
  - **56x** per-server efficiency improvement