Tesi di Laurea Magistrale in Ingegneria Informatica

Virtualized FPGA-acceleration of Distributed File Systems with SR-IOV and Containerizartion

Anno Accademico 2023/24

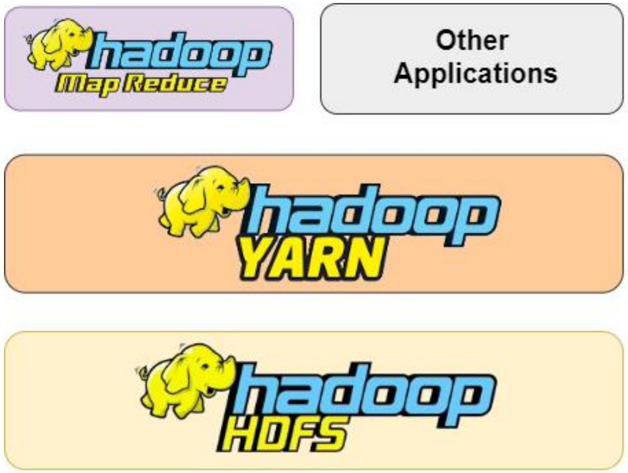
relatore
Ch.mo Prof Alessandro Cilardo

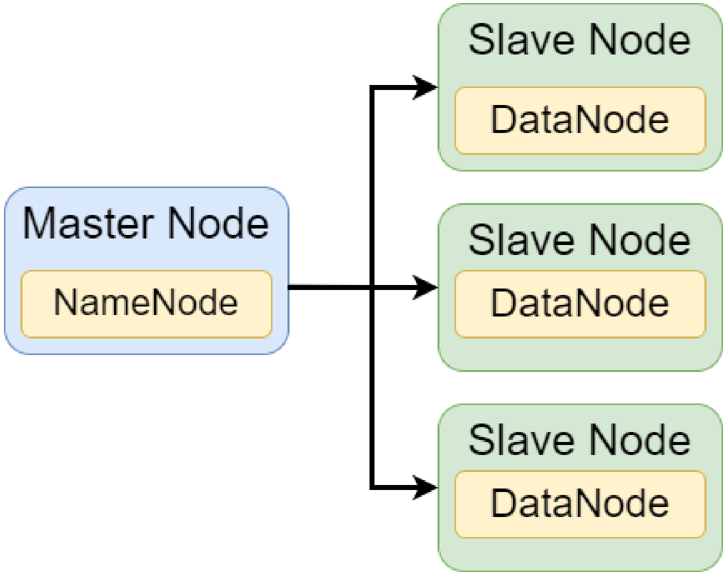
correlatore
Ing. Vincenzo Maisto

candidato
Biagio Spena
Matr. M63001373

What are Hadoop and HDFS ?

- Framework for distributed processing of datasets
- **Hadoop Distributed File System** (HDFS)
  - Stores user data
  - **Master-Slave Architecture**

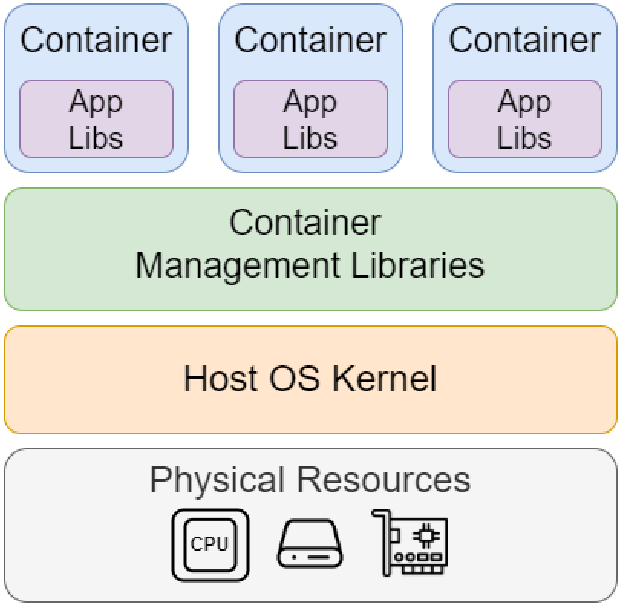




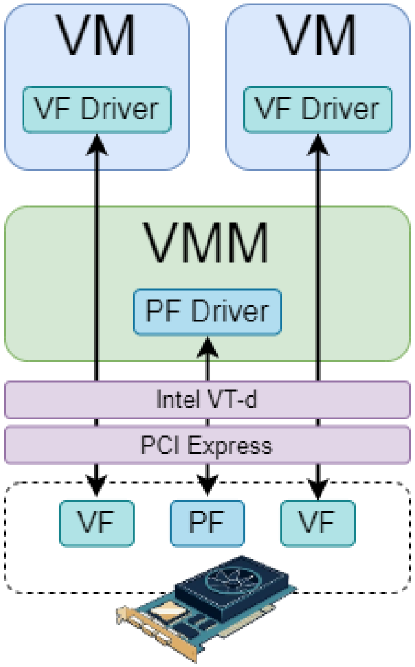
- **NameNode** (master daemon)
  - Stores metadata
- **DataNode** (slave daemon)
  - Stores the actual data

Virtualizing an Hadoop Cluster

- **Container Virtualization**
  - Lightweight virtualization
  - A container is an **isolated computing environment**

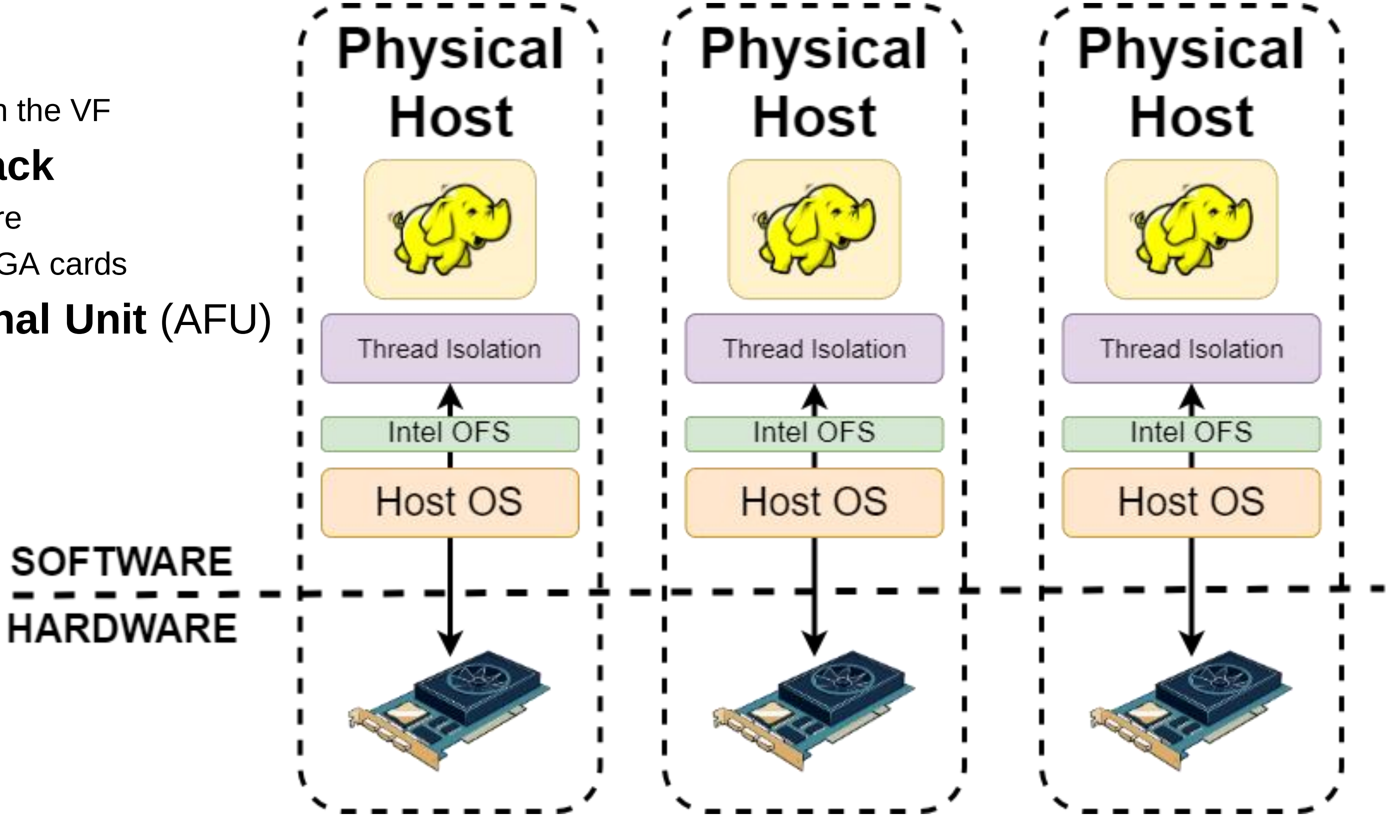


- **Single Root I/O Virtualization (SR-IOV)**
  - Share a single PCIe device across multiple VMs
  - **Physical Functions** (PF)
  - **Virtual Functions** (VF)
    - One or more VFs are directly assigned to a VM



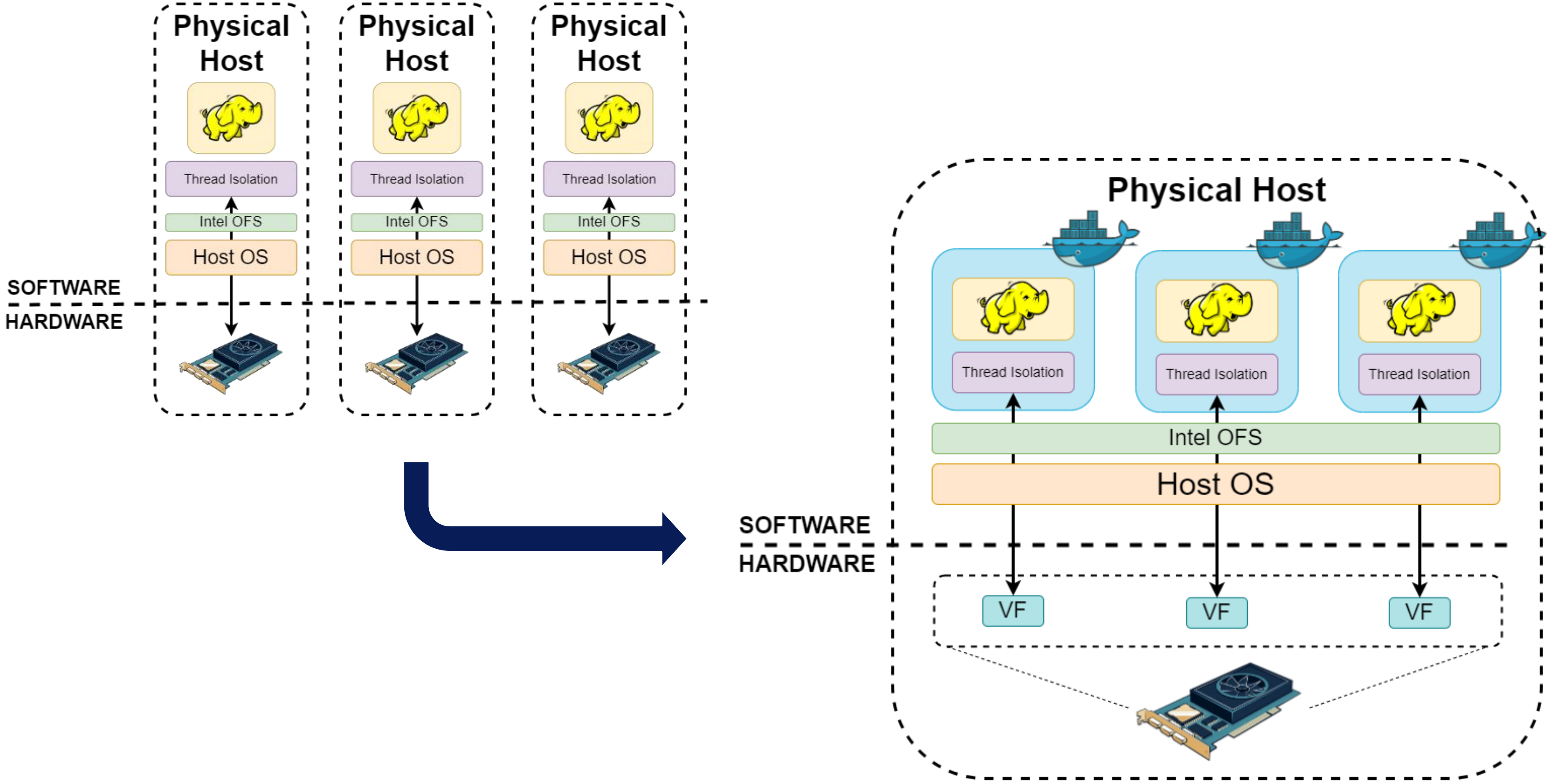
Starting Architecture: FPGA-Acceleration for HDFS

- **FPGA-accelerated Hadoop Daemons**
  - For Erasure Coding
- **Thread Isolation**
  - Interconnect HDFS with the VF
- **Intel Open FPGA Stack**
  - Stack hardware/software
  - Intel PCIe-attached FPGA cards
- **Accelerator Functional Unit** (AFU)
  - FPGA accelerator core
  - Exposed through VF
  - One VF is one AFU



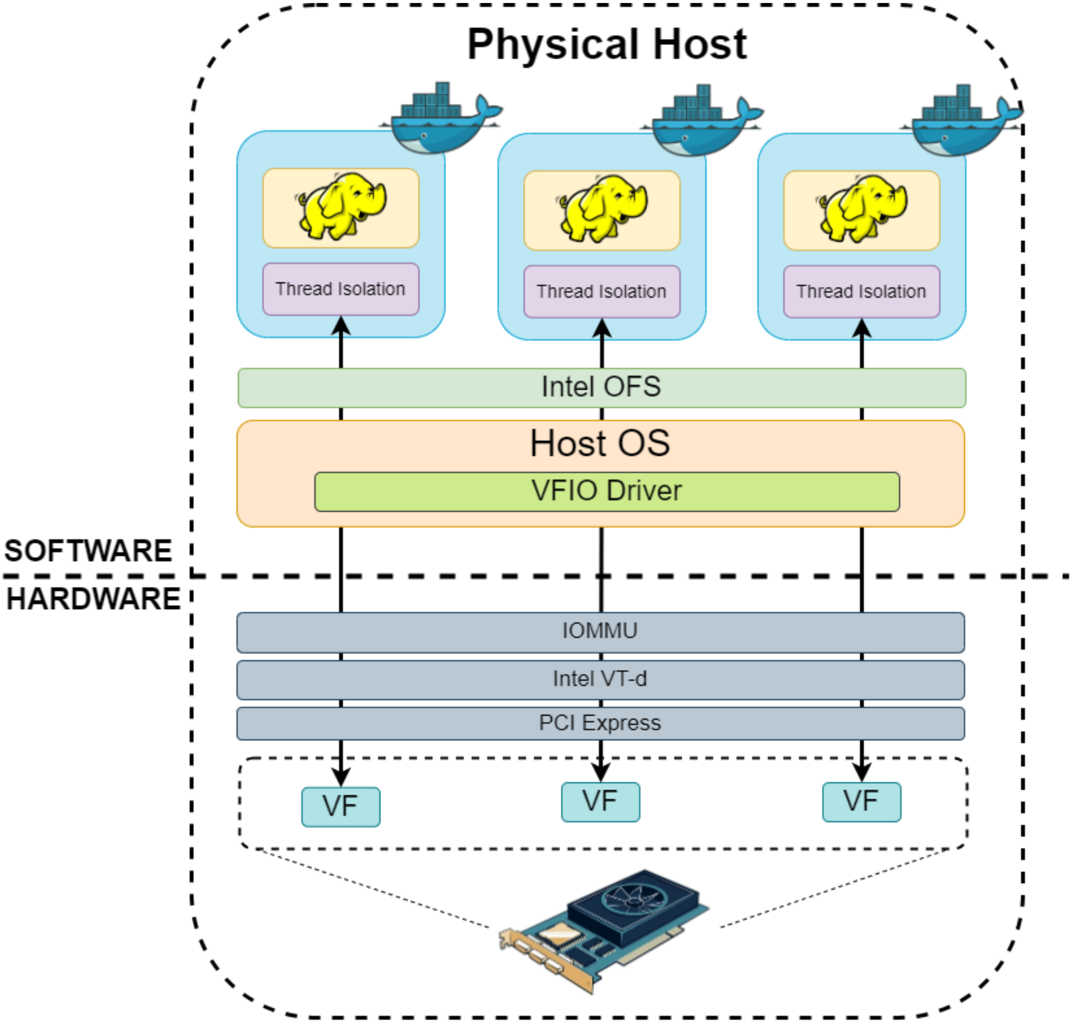
Thesis Objective

FPGA Acceleration from Physical Cluster to Virtual Cluster



Virtual Accelerated Cluster Architecture

- **Docker Container**
  - FPGA-accelerated Hadoop Daemons
  - Thread Isolation
- Intel OFS
- **VFIO Driver**
  - Exposes direct device access to userspace
- **Intel VT-d**
- **IOMMU-mapping**
  - Direct access to the host memory space from the device



Experimental Validation

- Validate the virtual cluster functionality and performance
  - **DFSIO** benchmark ⟶ Read/write on DFS
- Experimental Setup
  - **FPGA Acceleration** ⟶ **RS[3:2]**
  - FPGA device : **Max 13 VF**
  - **6 Containers**: 1 Master + 5 Slaves
  - 6 GB of RAM for each container

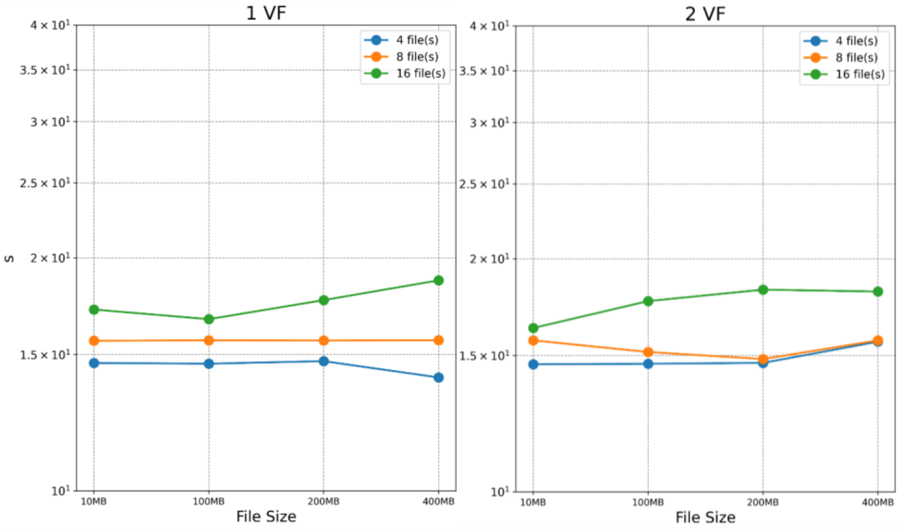
Indipendent Factors	Levels
Number of Files	4 8 16
Size of file	10 MB 100 MB 200 MB 400 MB
Number of VF per container	1 2

Response Variables for read/write	Description
Throughput (MB/s)	MB transmitted per second
Runtime (s)	Total operation execution time
Average I/O Rate (MB/s)	Average DFS read/write speed

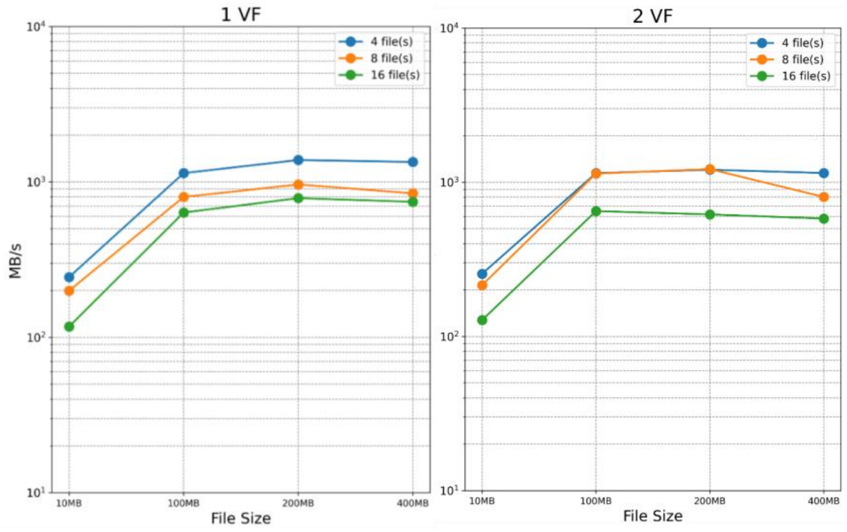
DFS Read Performance

- Read operations do not trigger erasure coding
  - ⟶ No interaction with VFs

Runtime

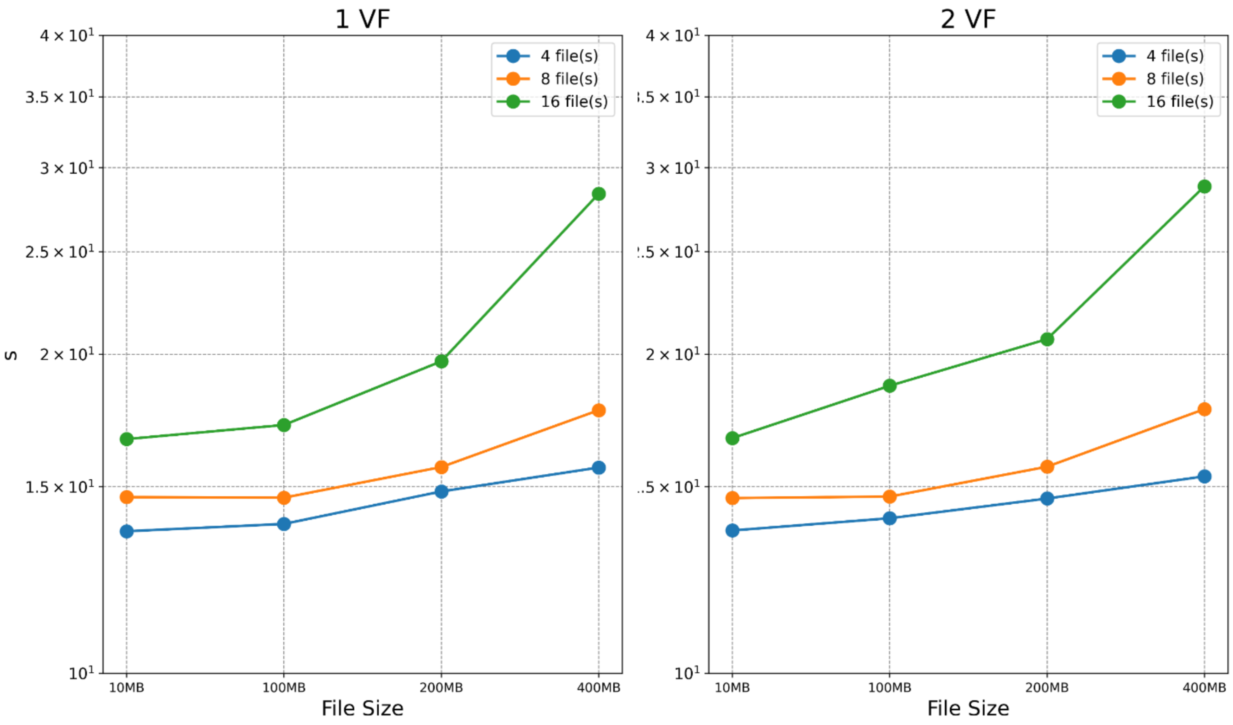


Throughput



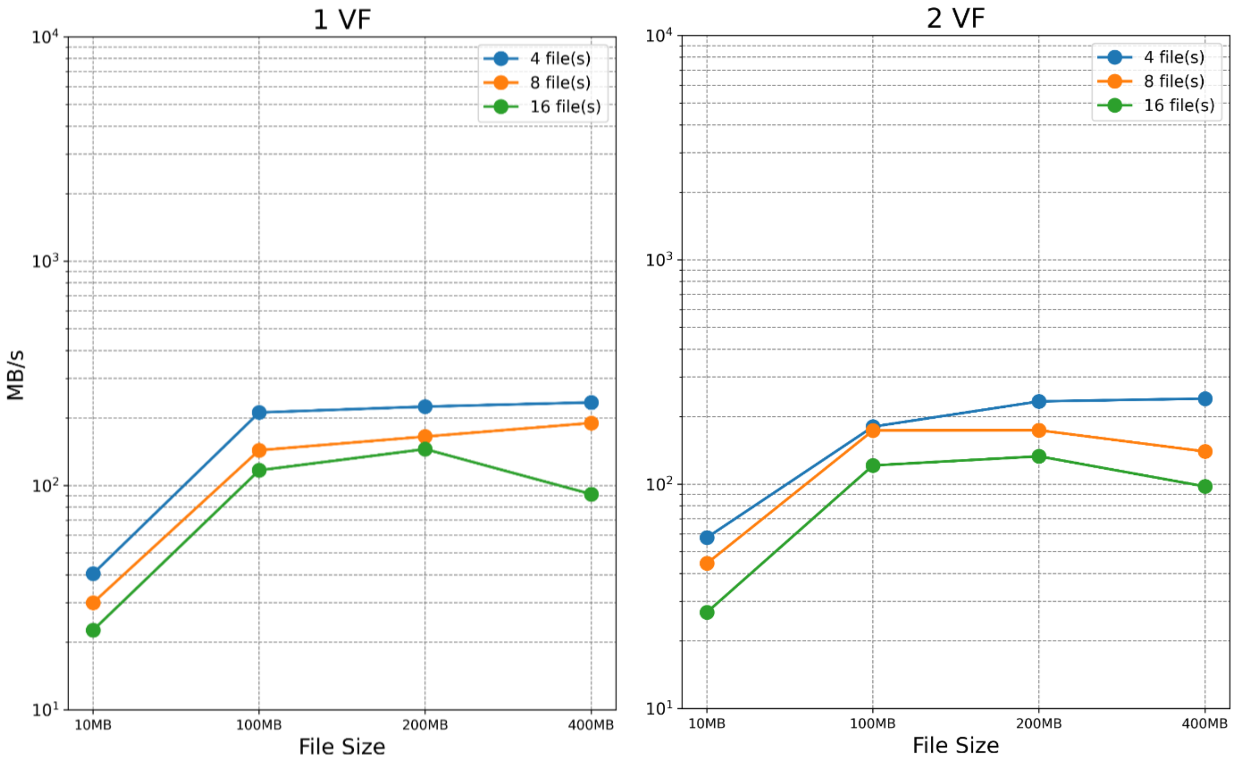
DFS Write Performance

- Runtime
  - No significant variations up to 8 files
  - Slight degradation for 16 files
  - 2 VFs show higher overhead



DFS Write Performance

- Throughput / IO rate
  - Comparable performance
  - 2 VFs better load handling for larger files
  - Large file sizes (>400 MB) difficult to handle



Conclusions

FPGA-acceleration of Virtual HDFS Cluster 👍

Physical Cluster

- One FPGA card per node
- More complex development
- Longer deploy time
- Lower portability
- Higher costs 👎
  - More components
  - More maintenance
  - Higher energy demand

Virtual Cluster

- One FPGA card per virtual cluster
- Easier and faster development
- Performance bottle-neck:
  - Host physical resources
- **Functionally equivalent!** 👍
- Reduced costs
  - Less components
  - Eaiser maintenance
  - More energy-efficient