Whamcloud

Cloud Oriented
Massive Data Movement and Management

Lei FENG (flei@ddn.com)

Oct 2020

DDN® STORAGE

Outline

- Background
- Challenges
- Optimization
- mpiFileUtils
- Use Cases

Background

Whamcloud

- Cloud Storage becomes more and more popular
  - Convenient
  - Capacity
  - Cheaper
- Local Storage is still necessary
  - Performance
  - Existing applications
  - Security Issue
- New Requirements
  - Copy/Move/Sync between local and cloud storage

Challenges - Massive Dataset

Whamcloud

- Capacity
  - 5TB / disk
  - 20 disks / node
  - 10 nodes / cluster
  - Totally: 1PB
- Number of files
  - Billions of files
- Huge File
  - TB level

Challenges - Cloud Storage (S3)

Whamcloud

- Simple model
  - No directory structure (actually a key-value storage)
- Weak functions
  - Object must be put as whole
  - Can not append data or overwrite part of data
  - Different attributes (stat/xattr) support
- Long latency
  - Latency: 100ms
  - TPS: 10
- High concurrency

Optimization - Parallelization

Whamcloud

- Multi-processes/threads on multi-nodes
  - Distributed file system can be mounted on multiple nodes
  - Compensate the long latency of S3 storage
  - Fully usage of CPU/IO/Network of multiple nodes
- Task split
  - Split by namespace
  - Split by chunk

Optimization – Change Monitoring

Whamcloud

- Traditional way
  - Initial start rsync
  - Monitor change lively
    - Inotify
      - Single node
      - Transient
  - Restart rsync
    - Scan both sides and compare
- Lustre ChangeLog
  - Series of events (timestamp, FID and operation)
  - Built-in Lustre file system and monitor changes from all clients
  - Persistent in MDT
- Rsync in Lustre way
  - Initial start rsync
  - Monitor: read ChangeLogs
  - Restart: read ChangeLogs
- No more scanning except initial copy!

mpiFileUtils – Overview

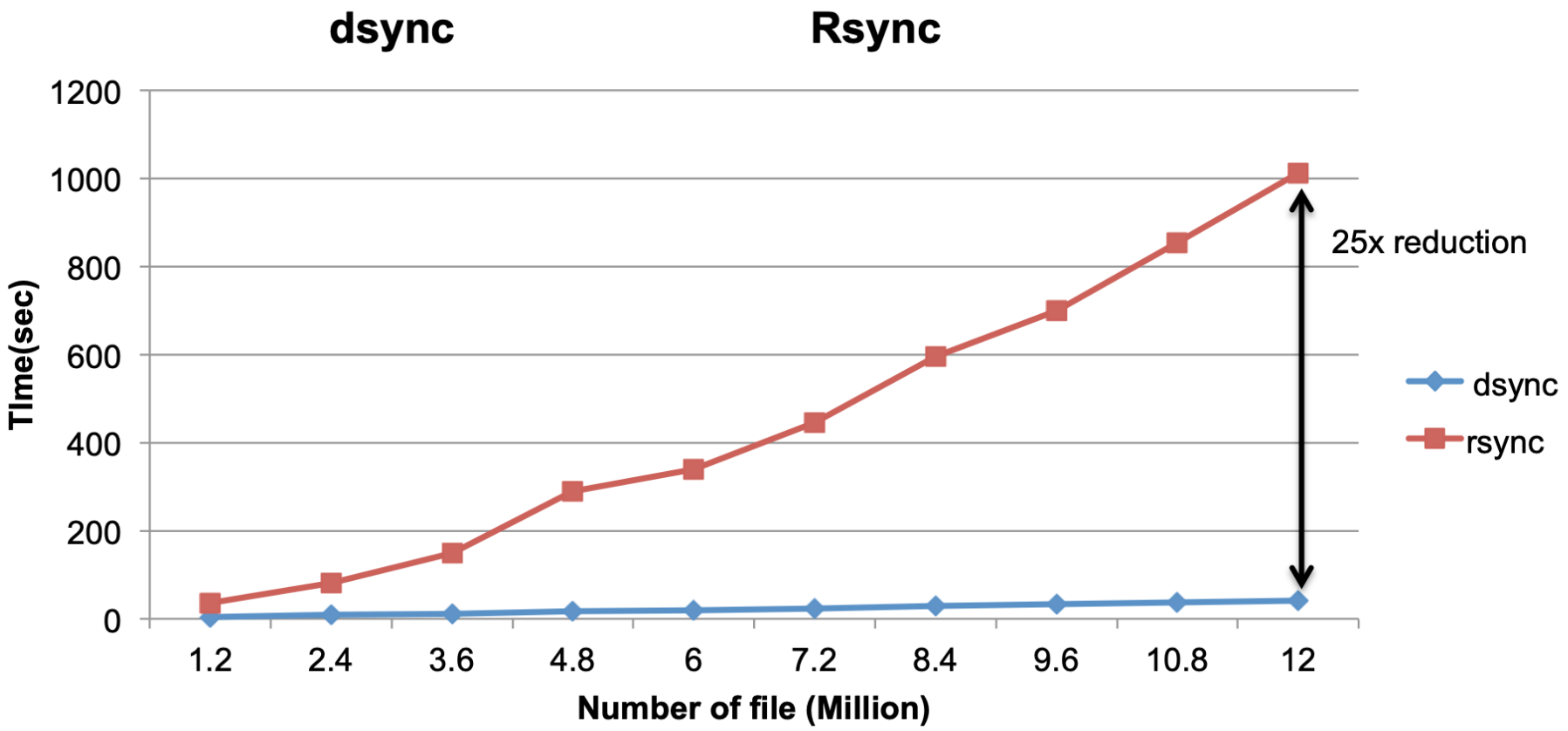
Whamcloud

- MPI (Message Passing Interface) based
- A Group of Utils
  - dcp, drm, dfind, dgrep, dsync…
- Large dataset: directory tree and/or file size
- Scalability
- Performance
- Compatibility

mpiFileUtils – Experimental performance

Whamcloud

dsync v.s. rsync

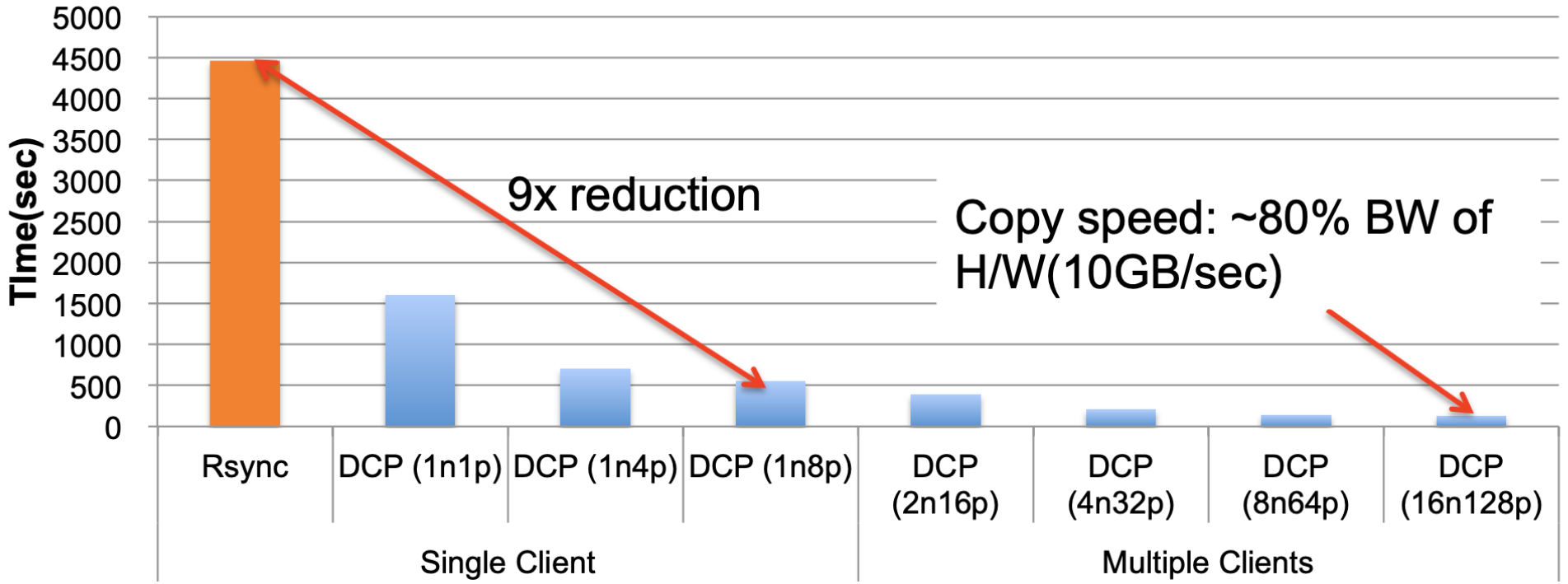


mpiFileUtils – Experimental scalability

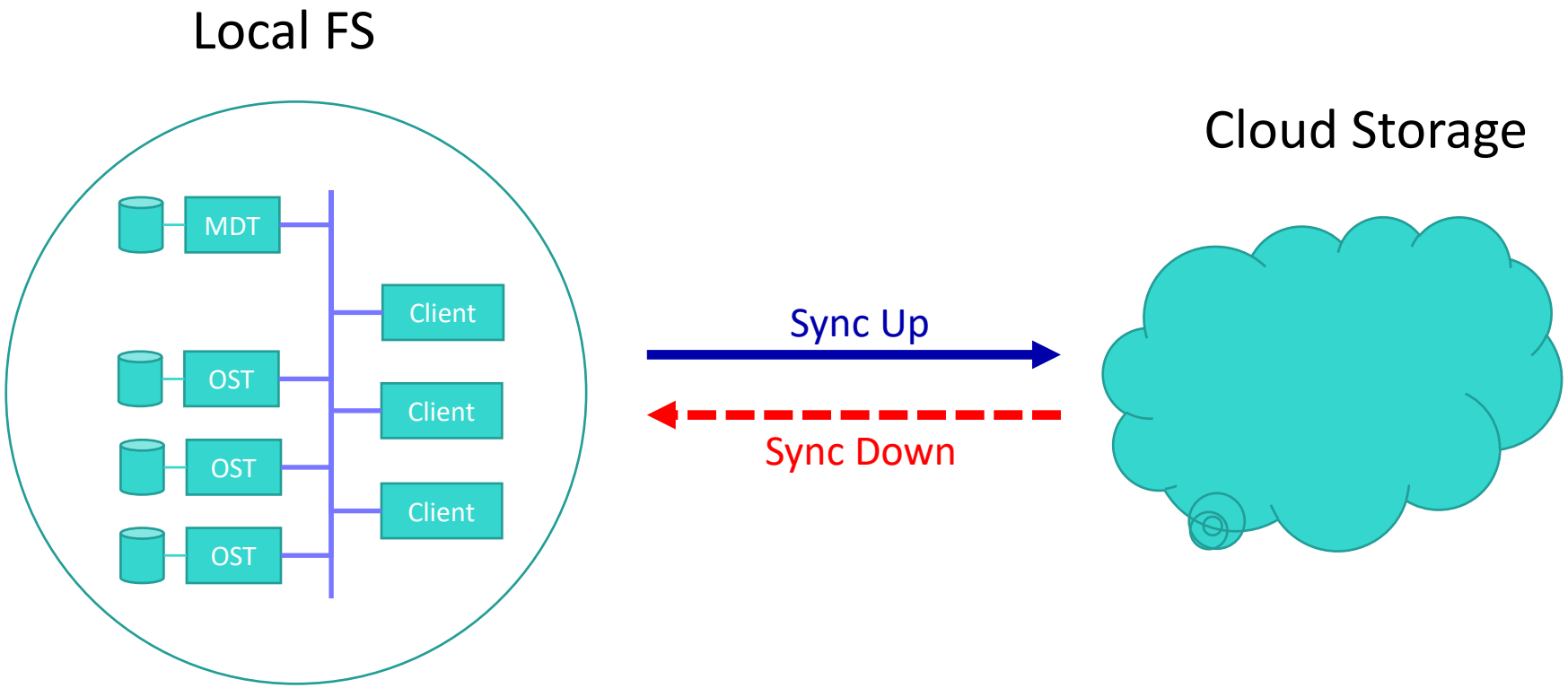
Whamcloud

dcp v.s. copy 1 TB single shared file

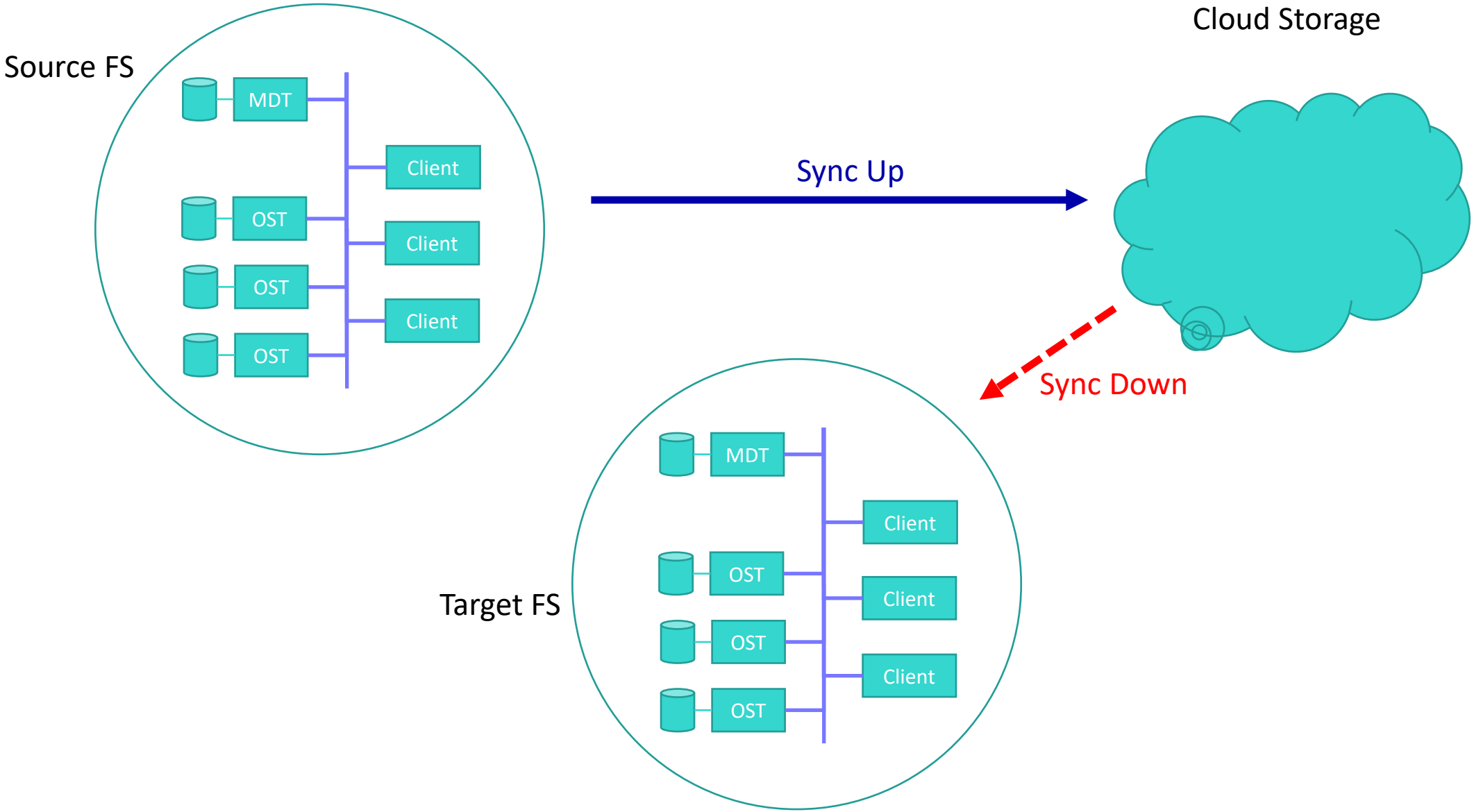
Primary System : 1 x DDN ES7K(140 x NLSAS), 2 x FDR
Backup System : 1 x DDN ES7K(140 x NLSAS), 2 x FDR



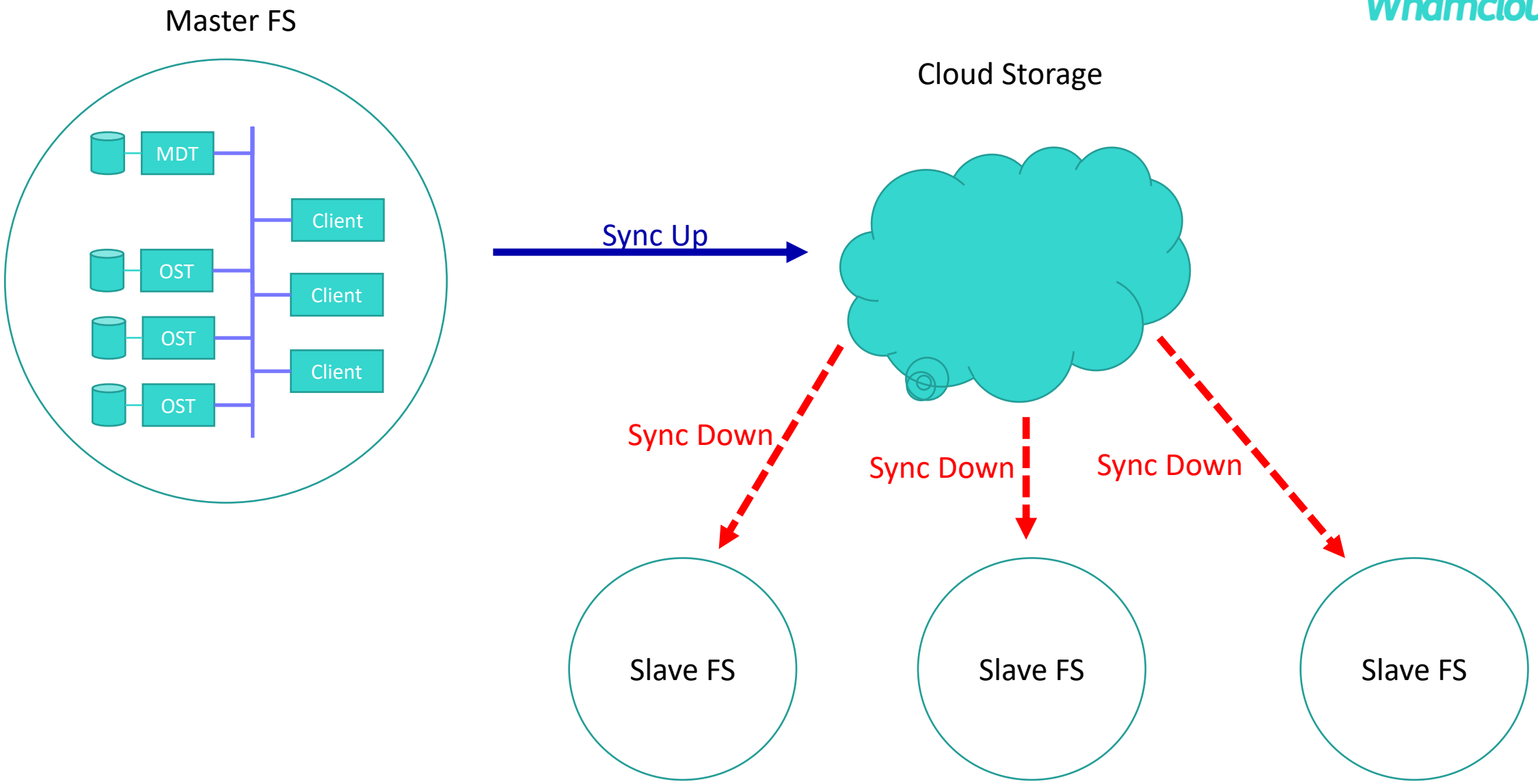
Use Cases – Backup



Use Cases – Data Migration



Use Cases – Data Sharing



Thanks!

Lei FENG (flei@ddn.com)