HPC storage solutions: overview and trends

Dr. Clement Onime

onime@ictp.it

Overview

- Background
- Storage solutions in HPC
- Recent trends

Clusters and Storage

- Common storage is a key shared resource which
  - Provides ability to execute computations on any node.
  - Reduces the copying/movement of data from one location to another.
- Scientific data may be expensive to re-create.

Storage and data

- Sharing
  - Mostly Quota based
- Data Lifecycle/lifespan
  - Generation/creation → re-use/analysis → deletion
  - For HPC applications, storage is just another memory device.
    - Necessary for capacity but expensive to consult in terms of CPU cycles
- Portable Operating System Interface (POSIX)
  - File operations
- Small Computer System Interface (SCSI)
  - Block operations.

Types of storage

- Local
  - Internal storage
- Remote
  - Network file-system/attached storage
  - Storage Area Network
  - Distributed file-systems

Local storage: Magnetic devices

- Magnetic
  - Basic design is magnetic storage devices
  - Ability to storage data for extended periods
  - Moving parts
  - Tapes
    - Sequential access, capacity: LTO-10 to offer 48TB
  - Hard-disk
    - Random access
    - Capacity
    - Data transfer rates
    - Speed (RPM)
    - SATA/SAS
    - Modern devices include RAM & Error Correction.

Images from : https://upload.wikimedia.org/ and https://en.wikipedia.org/

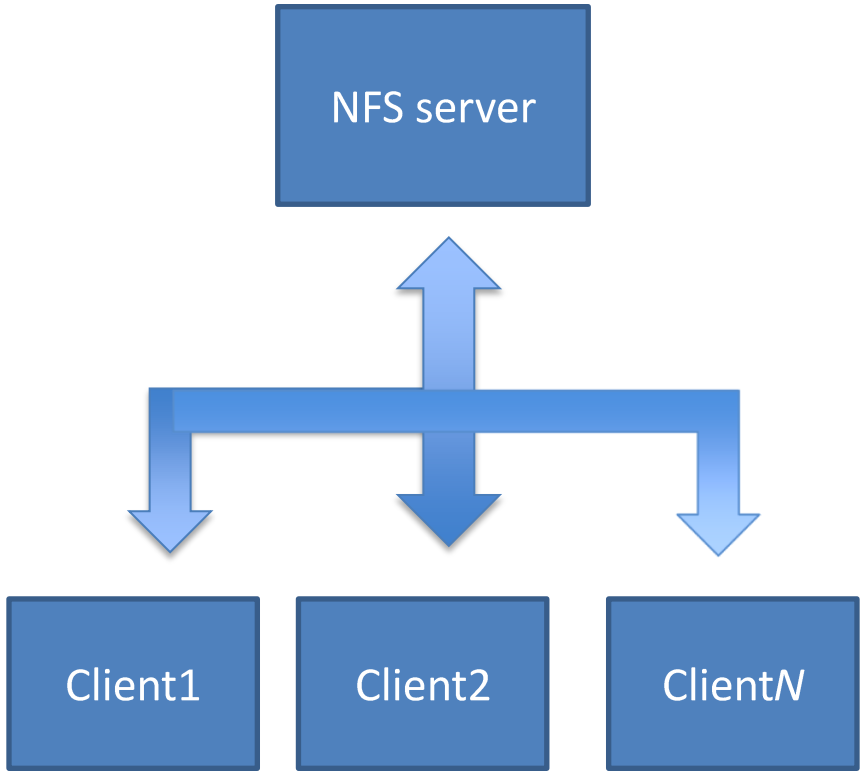
Local storage: Solid State Disks

- SSD devices
  - No moving parts
  - Random access devices
  - Increasing capacity
  - Underlying technology (NAND/DRAM)
  - New technology
    - Limited information about storage periods and reliability
  - Data transfer rates

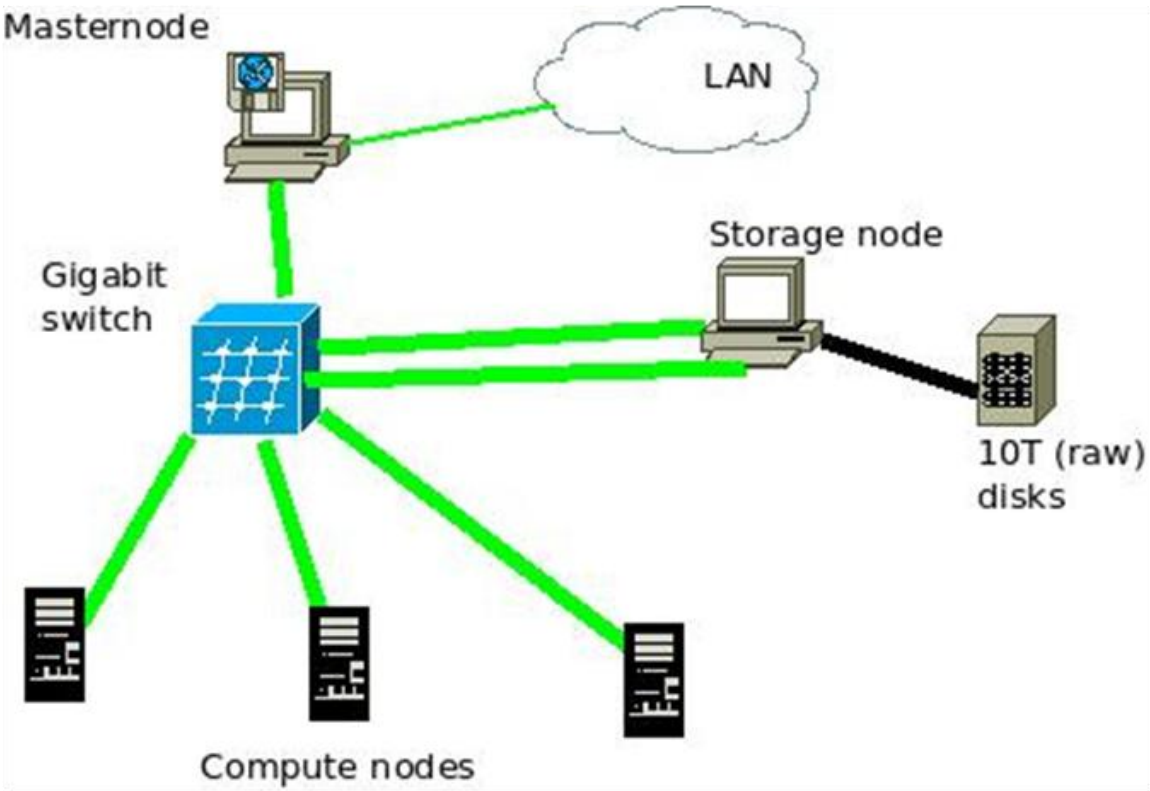
Images from: http://www.forensicmag.com/articles/2013/05/forensic-insight-solid-state-drives
And http://www.macworld.co.uk/news/mac/ssd-vs-hard-drives-which-best-storage-have-mac-3463640/

Remote: Network File Systems

- NFS protocol
  - Open standard
  - Client(s) ←→ Server model
    - simple
  - File level access over network
  - v4 tries to solve limitations in security and scalability
    - *(compare to CIFS & AFS)*

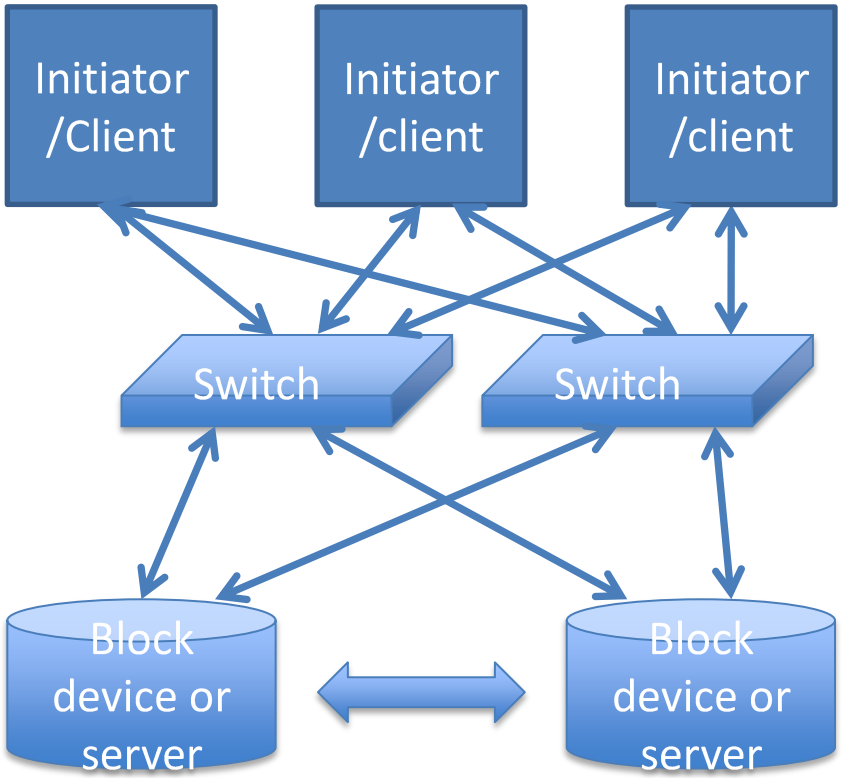


HPC storage (NFS) architecture



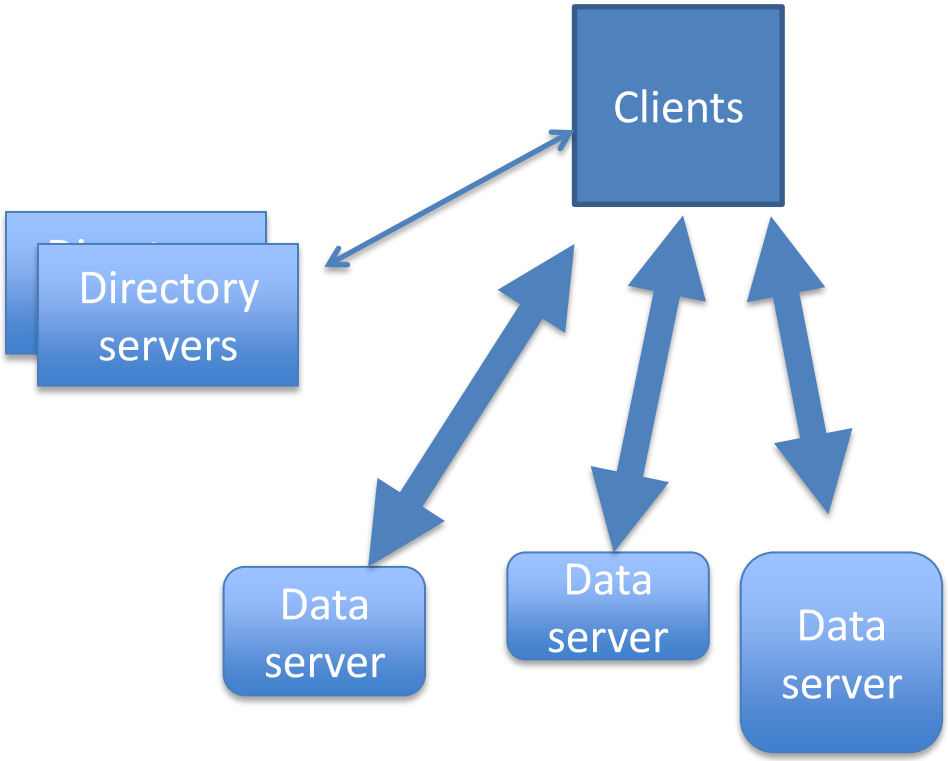
Remote: Storage Area Network

- Dedicated Network for storage related traffic
- Block Level Access
  - HBA fibre based solutions offer lower latency.
  - Ethernet: iSCSI
- Client ← → Server model
  - Failover/replication

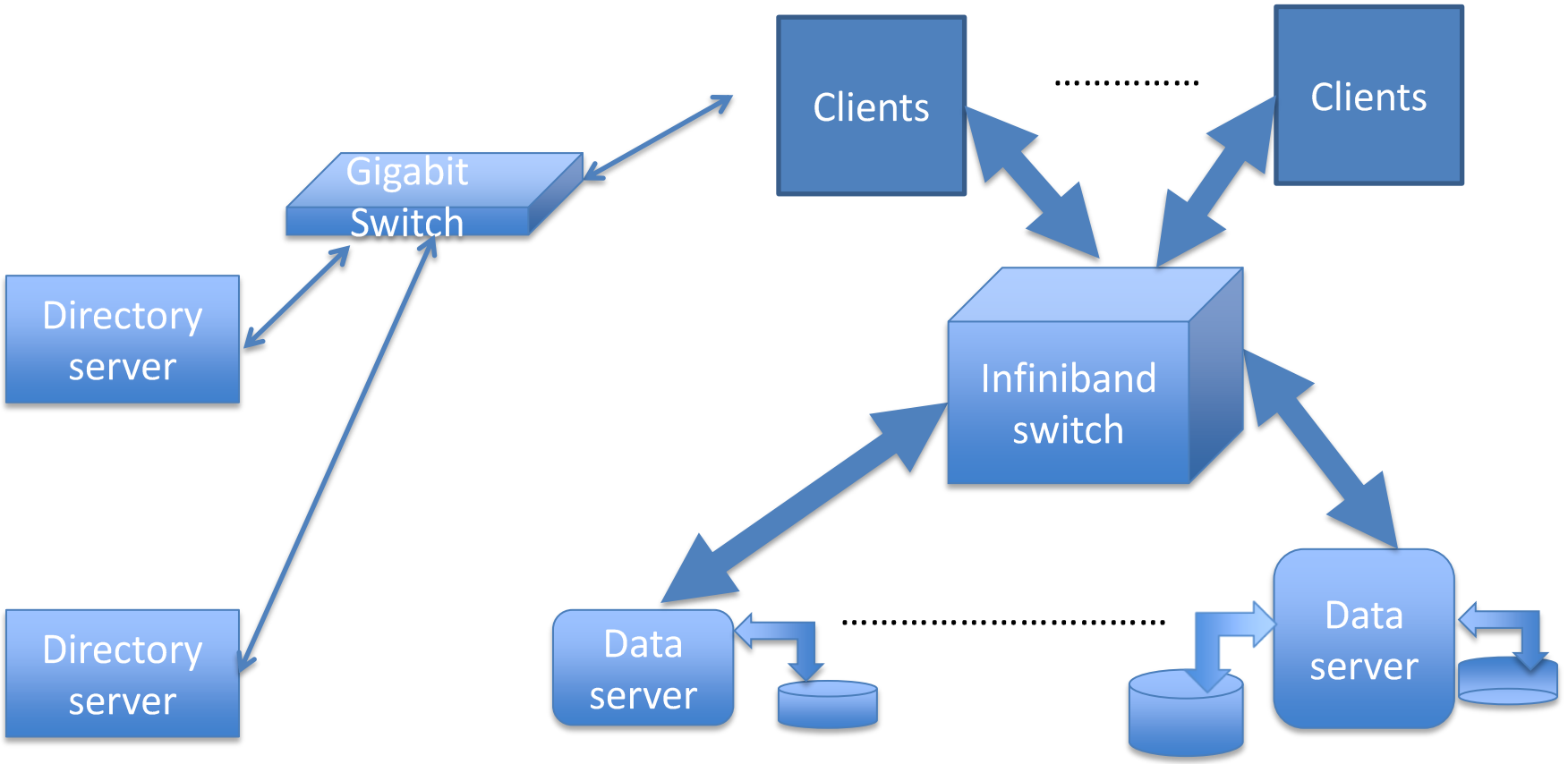


Remote: Distributed File System

- Single file system space is distributed.
- Client s← →Servers model
  - Independent Directory service separates metadata from data
  - Clients need dedicated software layer
- Examples: AFS, Lustre, parallel-NFS



Lustre HPC storage cluster



Hadoop

- Hybrid Framework for both
  - Data Storage
    - (HDFS)
  - Data processing
    - Batch processing
    - Resource manager
- Designed to prevent single points of failure

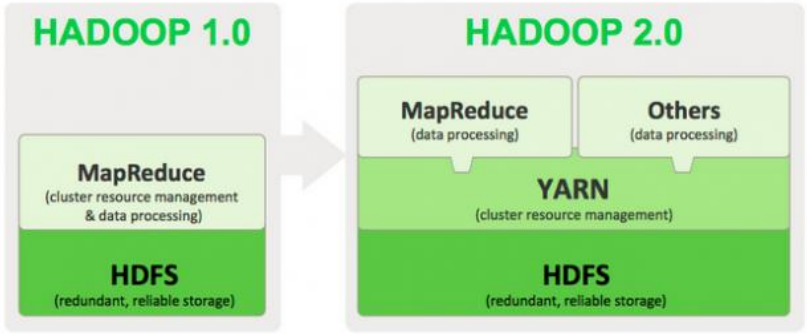


Image from http://opensource.com/life/14/8/intro-apache-hadoop-big-data

HDFS

- DFS approach
- Commodity hardware approach
  - *Separate servers not required*
- Features
  - Processing proximity
  - Data replication
  - *Not fully POSIX compliant*

HDFS Terminology

- Namenode
- Datanode
- DFS Client
- Files/Directories
- Replication
- Blocks
- Rack-awareness

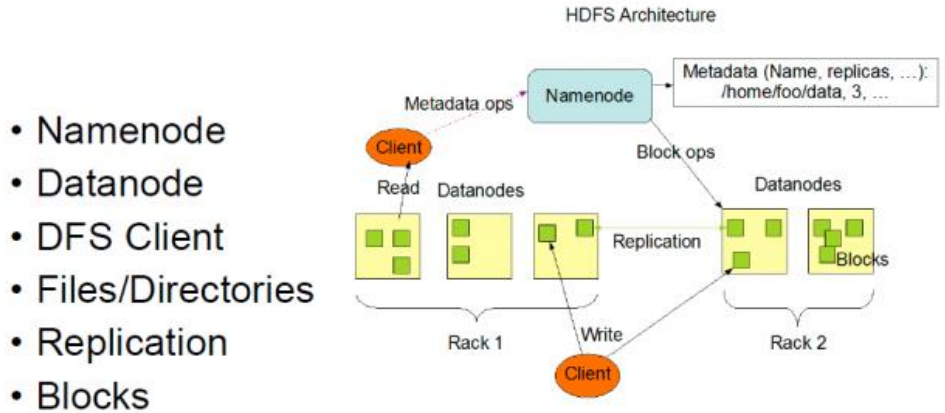


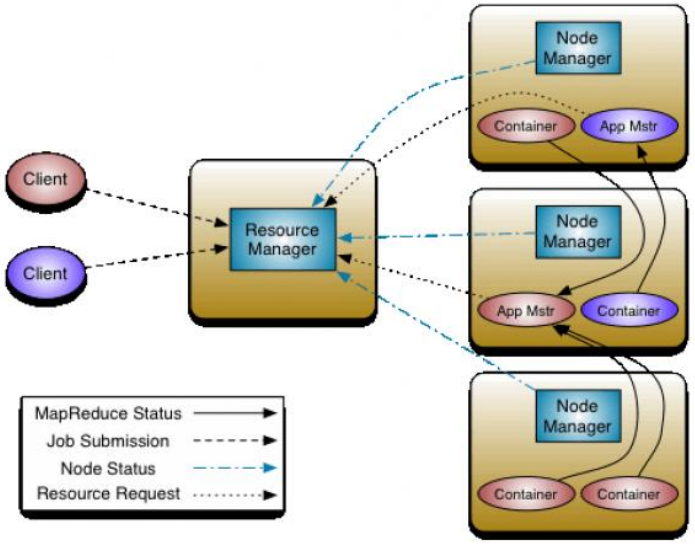
Image from http://opensource.com/life/14/8/intro-apache-hadoop-big-data

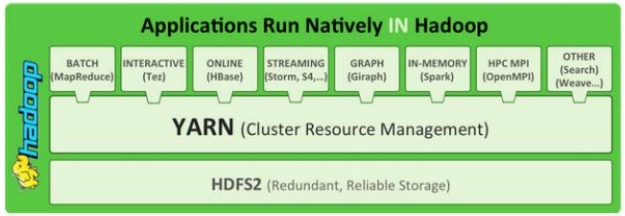
HDFS performance notes

- Namenode
  - Fast restart from failures due to secondary namenodes
- Datanodes
  - Block oriented DFS so files can be quite large
  - Replication/rebalancing
- Data awareness
  - Job tracker/scheduler
- Direct File access API in many languages
- Limitations
  - Poor concurrent writes
    - Throughput is high when dealing with immutable chunks
  - Jobs may require upload/download of data into HDFS
    - Could be long
  - No native O.S. clients
    - FUSE (Linux)

Hadoop Data processing

- Global Resource manager
- Node Manager
- Application Manager
- Container





Images from http://opensource.com/life/14/8/intro-apache-hadoop-big-data

Hadoop Ecosystem

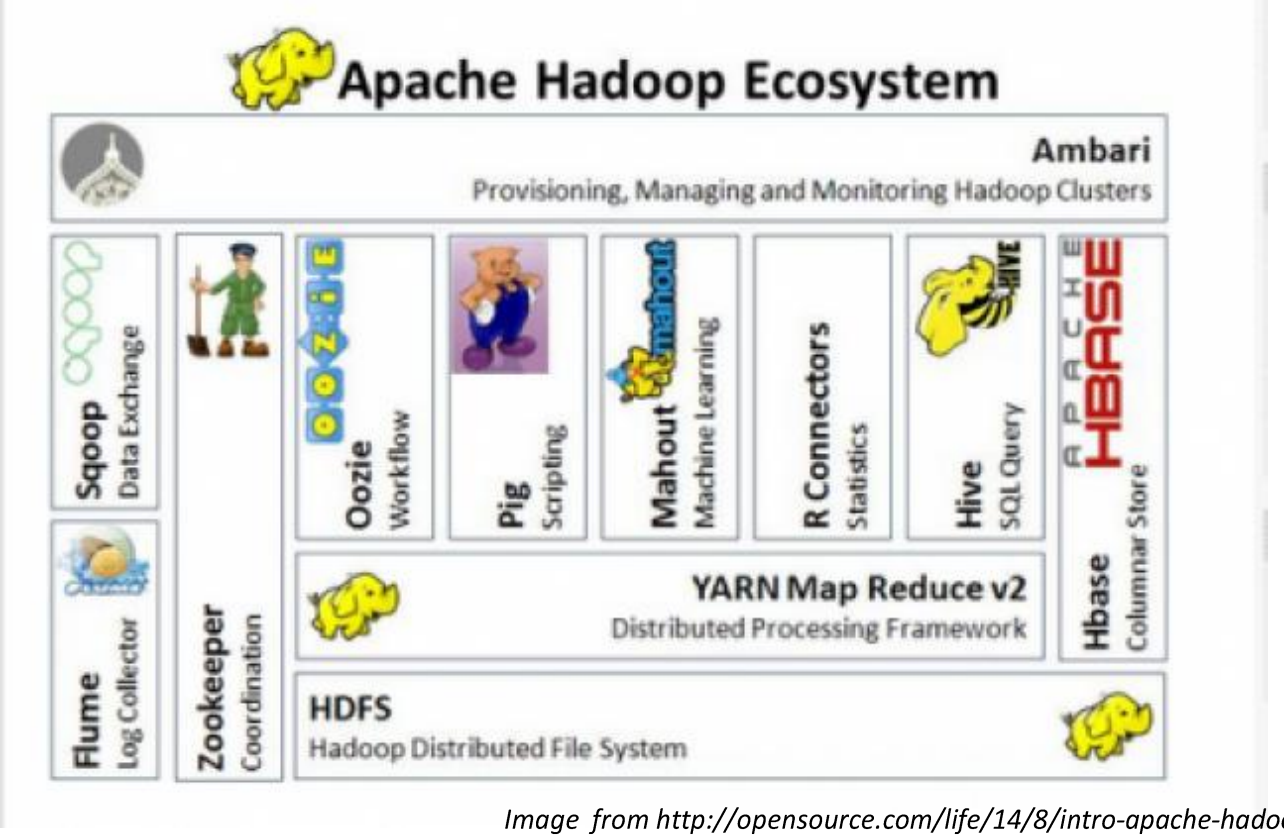


Image from http://opensource.com/life/14/8/intro-apache-hadoop-big-data

Summary

- Choosing the right storage solution for your HPC clusters is important for performance.
- It is possible to provide different storage solutions in a HPC cluster, however accessing same data seamlessly across solutions is not always possible.

Question-time

Thank you