

# Tehnologiile de comunicare intr-o infrastructura HPC

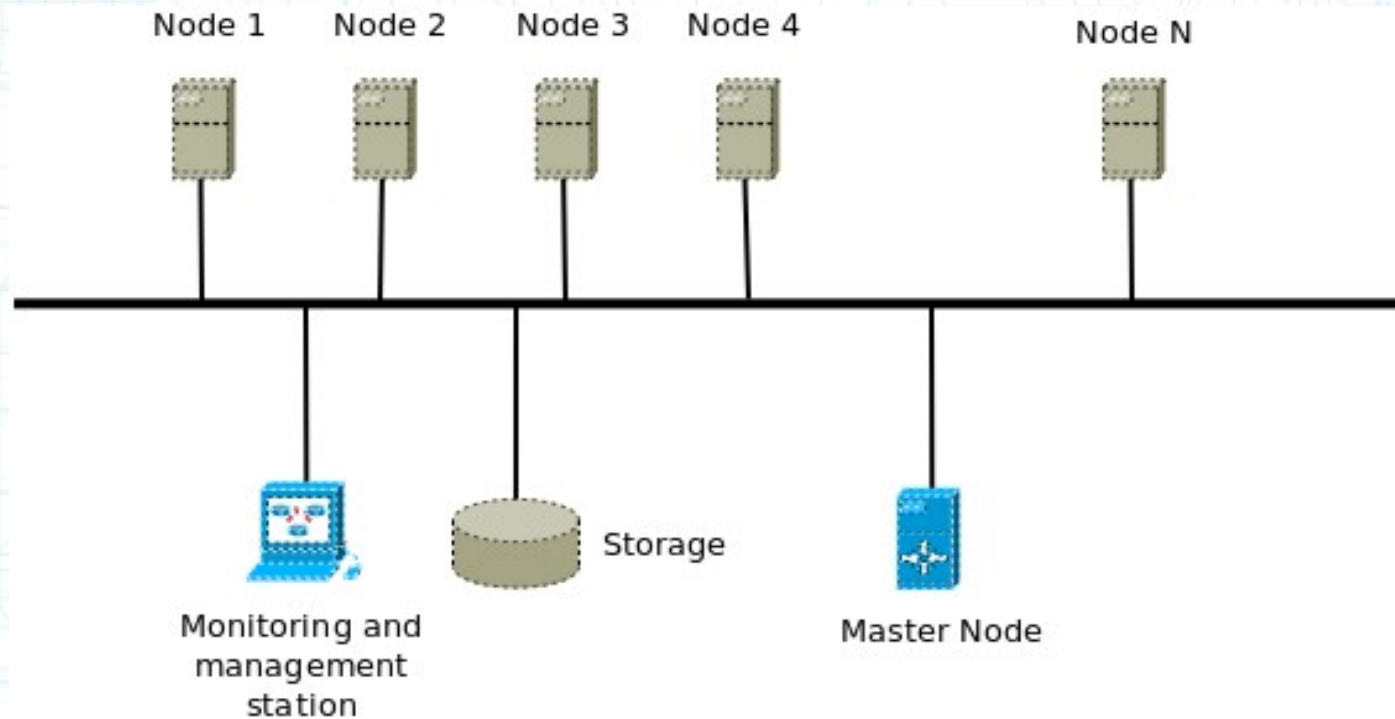
Alexandru Carp  
SRIC

# Tipuri de trafic

- Trafic computational
  - transfer de date intre noduri
- Trafic de stocare
  - intre noduri si infrastructura de “storage”
- Trafic de management
  - monitorizare, SSH, imagini de SO, configurari etc.
  - Trafic pentru job scheduling

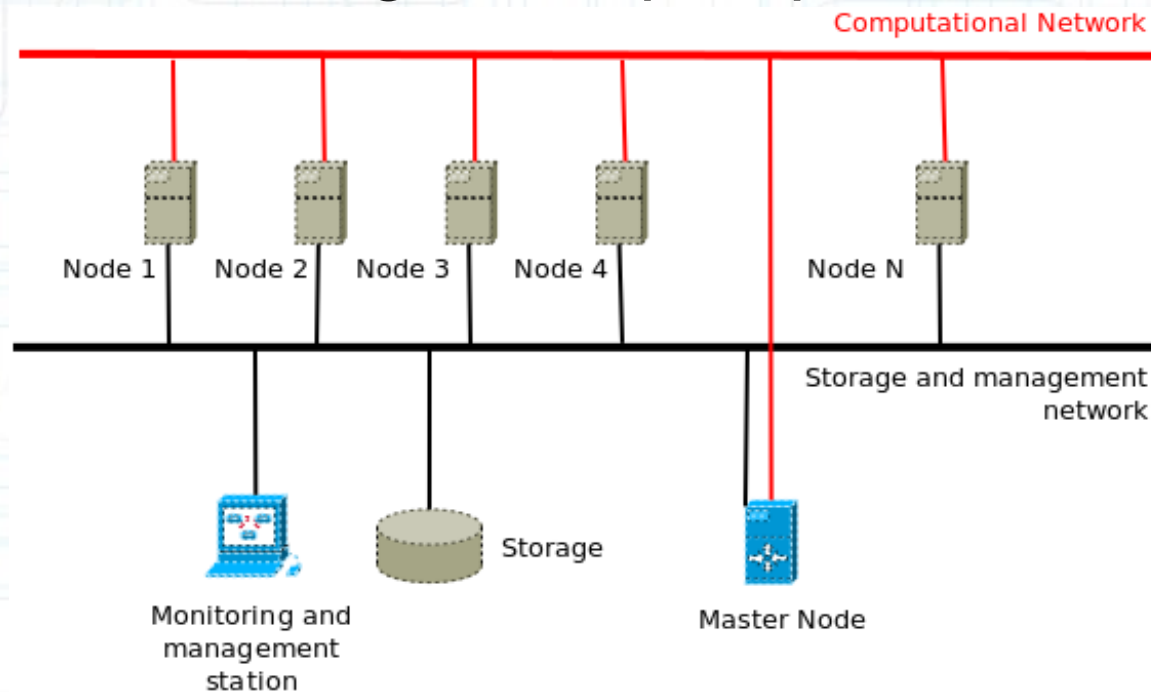
# Topologii posibile (1)

- O singura retea pentru toate tipurile de trafic
  - Constrangere: suport pentru TCP/IP



# Topologii posibile (2)

- O retea pentru traficul computational
- O alta retea pentru traficul de management + storage
  - Constrangere: suport pentru TCP/IP

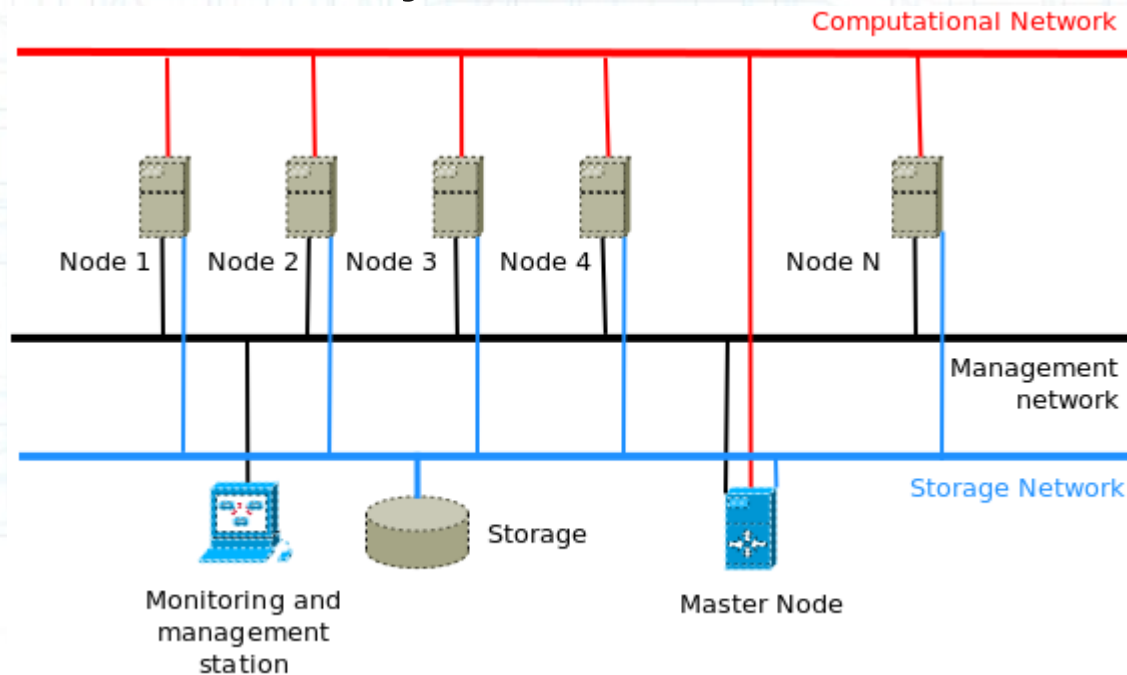


# Topologii posibile (3)

- O retea pentru traficul de management
- O alta retea pentru traficul computational si storage
  - Constrangere: Controlul congestiei intre traficul computational si cel de storage
  - Nu este potrivita cand avem aplicatii atat CPU intensive, cat si I/O intensive

# Topologii posibile (4)

- Cate o retea separata pentru fiecare tip de trafic
  - Avantaj: Resurse dedicate pentru fiecare tip de trafic
  - Dezavantaj: Costuri mari de intretinere



# Reteaua de management

- **Protocoale:**
  - SSH, SNMP, TFTP etc.
- **Particularitati ale traficului**
  - pachete mici
  - latime de banda mica
- **Tehnologii folosite**
  - Fast Ethernet (100Mbps)
  - Gigabit Ethernet (1Gbps)

# Reteaua computatională

- Cerințe
  - latime de bandă mare
  - latență mică
  - MTU mare
- Tehnologii folosite
  - Gigabit / 10 Gigabit Ethernet
  - Infiniband
  - Myrinet / Myrinet 10G
- Oversubscription
  - de obicei între 1:2 și 1:4



# Gigabit / 10 Gigabit Ethernet

- Mediu fizic
  - cupru sau fibra optica
  - 10 Gigabit Ethernet suporta si cabluri de tip Infiniband (tranzitie facila)
- Avantaje
  - Compatibilitate logica intre standarde (100Mbps / 1Gbps / 10Gbps)
  - latime de banda mare
  - retea convergenta => costuri mici
- Dezavantaje
  - latenta nesatisfacatoare in aplicatii critice

# Infiniband

- Latenta foarte mica (100 - 200ns)
  - arhitectura switched fabric (crossbar)
- Aplicatii
  - comunicatie interprocesor
  - comunicatie cu dispozitive de I/O
- Latime de banda
  - intre 2Gbps si 300Gbps
  - suporta agregarea legaturilor (1X, 4X, 12X)
  - tipuri de semnalizare: SDR, DDR, QDR, FDR, EDR

# Myrinet / Myrinet 10G

- Overhead mult redus fata de Ethernet
  - Thoughtput de 1.98Gbps pe o legatura fizica de 2Gbps
- Latenta mai buna
- Mediu optic
- Compatibil la nivel fizic cu 10 Gigabit Ethernet
- Popularitatea sa a inceput sa scada fata de Infiniband si Gigabit Ethernet

# Reteaua de stocare

- ATA over Ethernet
- Fibre Channel Protocol (SCSI over Fiber Channel)
  - Fibre Channel over Ethernet
  - Fibre Channel over IP
- iSCSI (SCSI over TCP/IP)
- iSCSI over Fibre Channel



# Concluzii

- Retele mici (zeci de noduri)
  - poate fi suficienta o singura retea (Gigabit sau 10 Gigabit Ethernet)
- Retele medii si mari (sute / mii de noduri)
  - in cazul aplicatiilor CPU intensive, este necesara o latentă mica pentru traficul computational => il vom separa intr-o retea diferita (Infiniband)
  - in cazuri I/O intensive, putem separa si traficul de stocare

# Intrebari

