

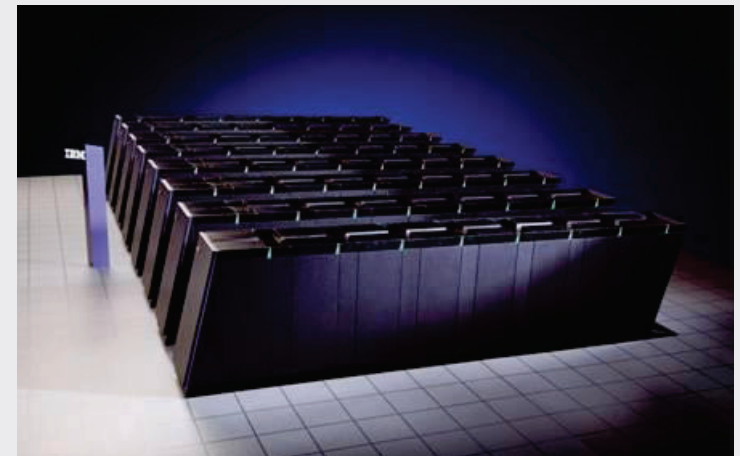


# Korištenje računalnih klastera

E. Imamagić, Srce  
Klasterska i grid radionica Srca

# Paralelno sklopovlje

- ❖ Superračunala
  - ♦ najmočnija dostupna računalna
  - ♦ za zahtevne računalne probleme
- ❖ Vektorska superračunala
  - ♦ Cray SV, NEC SX
  - ♦ Cray-1: 1976, 160.00 MFLOPS  
**8.8 M\$**
- ❖ Velika paralelna računalna
  - ♦ MPP (engl. *massively parallel processors*)
  - ♦ Connection Systems CM1 i CM2, Earth Simulator, Blue Gene
  - ♦ IBM – Blue Gene/L: 2005, 360.00 (70.72) TFLOPS, **~100 M\$**



# Računalni klasteri

- ❖ “... skup umreženih samostalnih računalov, ki predstavljajo enotni računalni resurs”  
R. Buyya
- ❖ Ekonomska alternativa za superračunala
  - ♦ sestoji se od standardnih komponent
  - ♦ “superračunalo za siromahe”
- ❖ Ključni faktori (80-tih)
  - ♦ mikroprocesori
  - ♦ brze mrežne tehnologije
  - ♦ Parallel Virtual Machine (PVM)



# Računalni klasteri

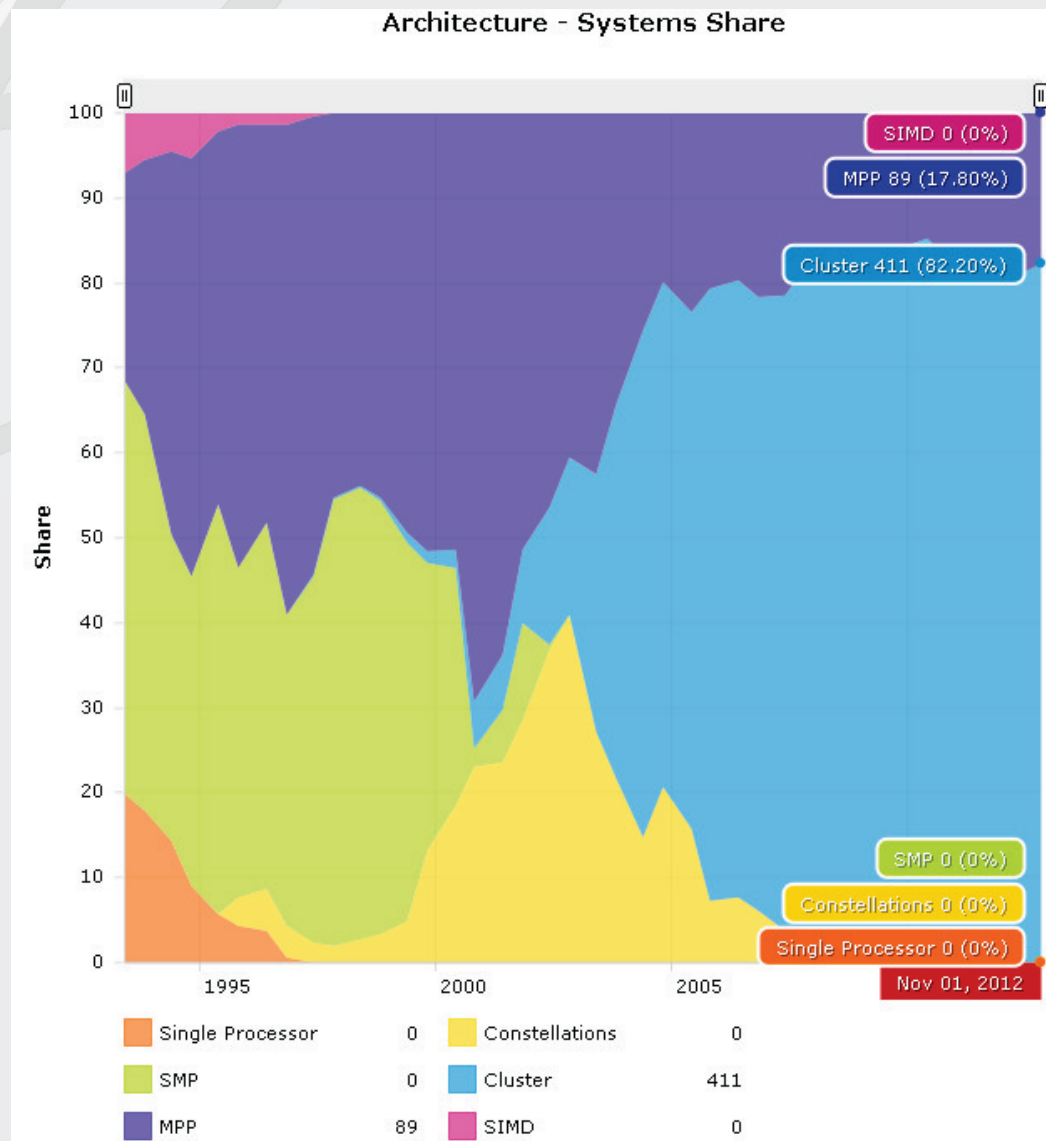
## ❖ Beowulf klaster (1994)

- ♦ T. Sterling & M. Baker
- ♦ NASA Ames Centre

## ❖ Danas

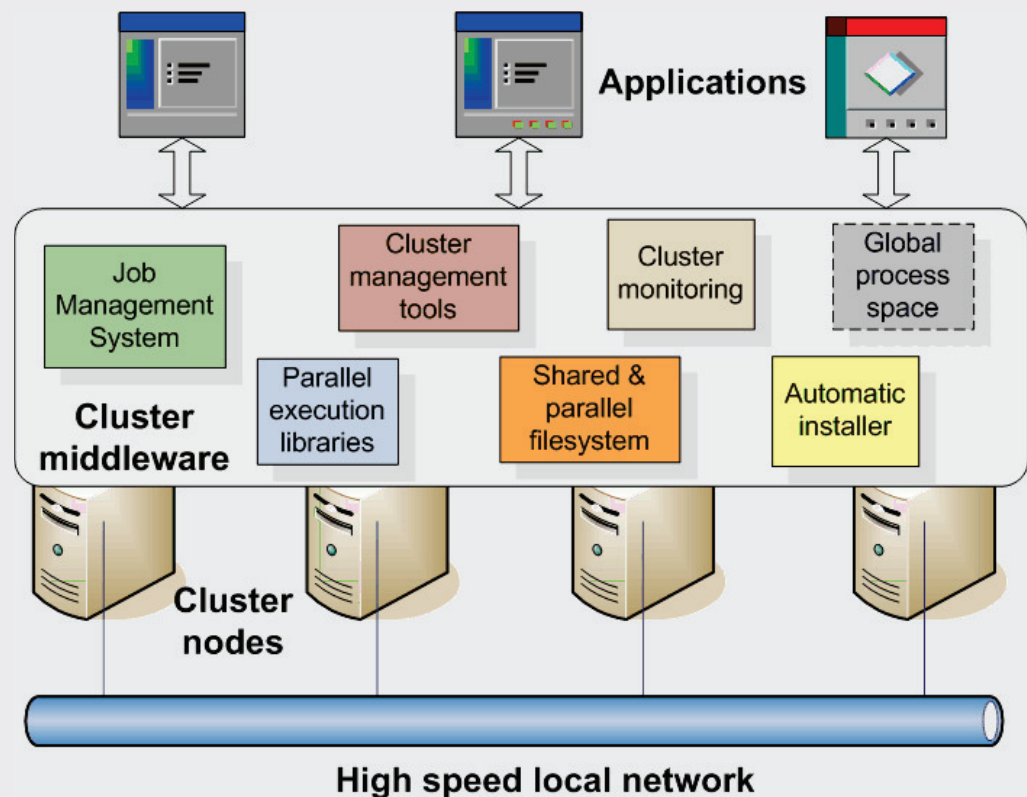
- ♦ široko rasprostranjeni u paralelnom računarstvu
- ♦ 82,2% Top500 liste najmoćnijih računala

# Računalni klasteri



# Arhitektura klastera

- ❖ Komponente klastera
  - ♦ računalni čvorovi
  - ♦ mreža
  - ♦ operacijski sustav (OS)
  - ♦ **klasterski posrednički sustav** (engl. *middleware*)

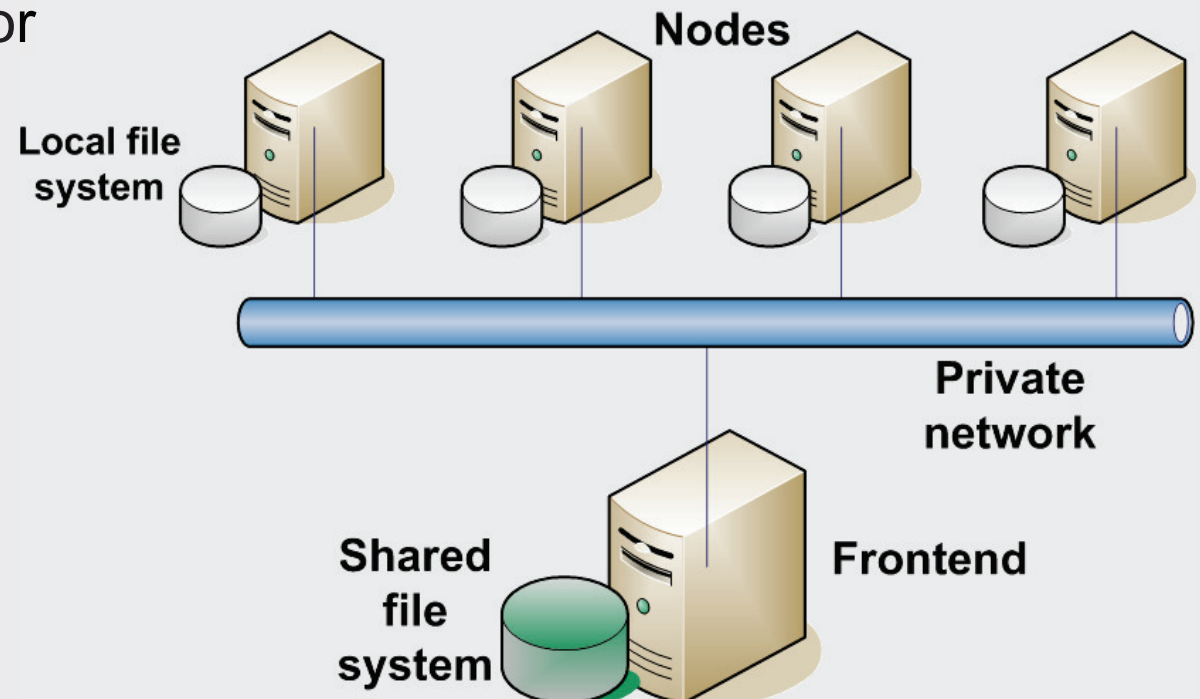




# Beowulf arhitektura

## ❖ Frontend

- ♦ pristupno računalo
- ♦ pristup pomoću protokola ssh
- ♦ središnji servisi klasterских posredničkih sustava
- ♦ dijeljeni diskovni prostor (direktorij **/home**)



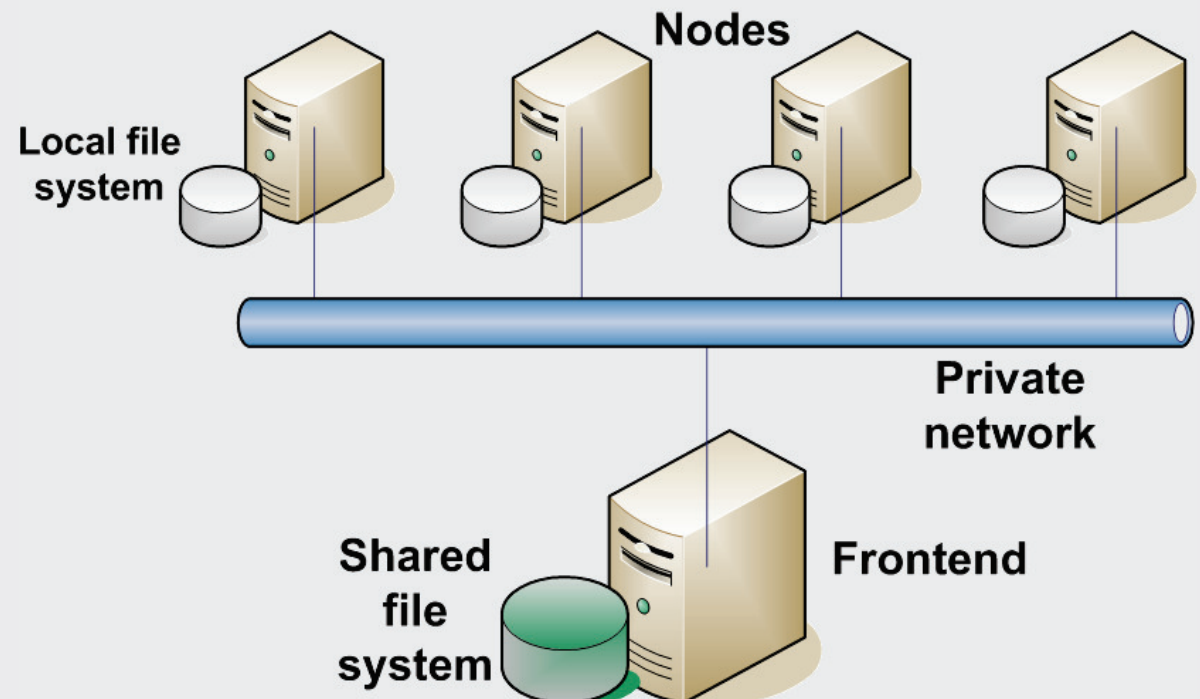
# Beowulf arhitektura

## ❖ Čvorovi

- ♦ dedikirana računala za izvođenje poslova
- ♦ povezani privatnom mrežom
- ♦ lokalni diskovni prostor (**/scratch**)

## ❖ Privatne mreže

- ♦ visoka propustnost
- ♦ malo kašnjenje
- ♦ moguće korištenje više mreža





# Komponente klastera

## ❖ Računalni čvorovi

- ♦ poslužitelji, radne stanice, osobna računala
- ♦ 32-bitna i 64-bitna arhitektura procesora
  - Intel Xeon, Itanium, AMD Opteron, Alpha

## ❖ Mrežna tehnologije

- ♦ Gigabit Ethernet, Infiniband, Myrinet, Quadrics QsNet II

## ❖ OS

- ♦ Linux, HP-UX, Solaris, MS Windows

# Sustavi za upravljanje poslovima

- ❖ Upravlja izvođenjem korisničkih aplikacija na klasteru
  - ♦ engl. *Job Management System* (JMS), *batch system*
- ❖ Posao – izvođenje aplikacije
- ❖ Vrste poslova
  - ♦ batch – ne zahtijevaju korisničku interakciju (npr. dugotrajna računanja, obrada podataka)
  - ♦ interaktivni – zahtijevaju korisničku interakciju (npr. grafička sučelja, unos dodatnih podataka, potvrđivanje akcija)
  - ♦ serijski – zahtijevaju samo jedan procesor za izvođenje
  - ♦ paralelni – zahtijevaju više procesora za izvođenje

# Sustavi za upravljanje poslovima

- ❖ Korisničko sučelje za podnošenje i upravljanje poslovima
  - ♦ podnošenje, upravljanje, dohvat detaljnih informacija o izvođenju
- ❖ Omogućava definiranje politike korištenja klastera
  - ♦ npr. *fair-share*, rezervacije, prioriteti, ...
- ❖ Optimizira korištenje čvorova
  - ♦ raspoređivanje poslova po čvorovima
  - ♦ praćenje opterećenja čvorovova

# Sustavi za upravljanje poslovima

## ❖ Otvoreni problemi

- ♦ spremanje stanja poslova (engl. *checkpointing*)
- ♦ selidba poslova između čvorova (engl. *process migration*)
- ♦ *preemption*
- ♦ rezervacija čvorova (engl. *advance reservation*)
- ♦ otpornost na greške u radu čvorova (engl. *fault tolerance*)

# Sustavi za upravljanje poslovima

## ❖ Son of Grid Engine (SGE)

- ♦ napredno raspoređivanje poslova (*fair share, tickets*)
- ♦ definiranje projekata i grupa korisnika
- ♦ detaljno opisivanje čvorova
- ♦ prilagodba okoline za specifične paralelne aplikacije

## ❖ Torque + Maui

- ♦ napredno raspoređivanje poslova
- ♦ rezervacije resursa
- ♦ precizno definiranje resursa kod paralelnih poslova

# Klaster Isabella

## ❖ Računalni klaster na Srcu

- ♦ dostupan cijeloj akademskoj zajednici
- ♦ <http://isabella.srce.hr>

## ❖ Arhitektura

- ♦ brza mreža Infiniband
- ♦ Son of Grid Engine
- ♦ dijeljeni datotečni sustav FhGFS
  - koristi Infiniband
- ♦ visoka dostupnost frontenda i podatkovnih elementa



isabella  
CLUSTER



# Klaster Isabella

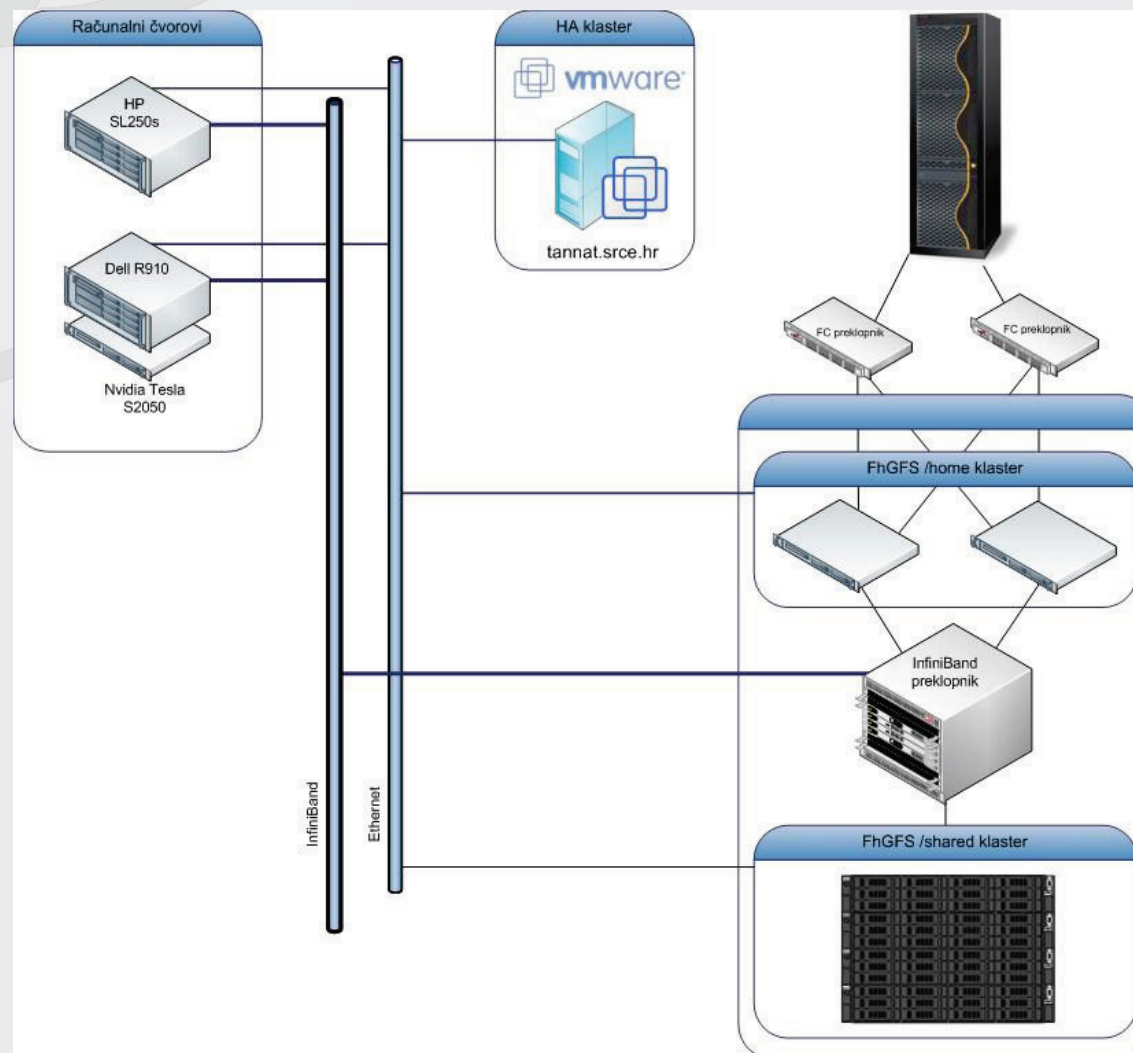
## ❖ Brojke

- ♦ 28 računalnih čvorova
- ♦ 480 procesorskih jezgri
- ♦ 4 grafička procesora (GPU)
- ♦ 2 TB RAM
- ♦ 4 TB dijeljenog diskovnog prostora
- ♦ 36 TB dijeljenog privremenog prostora (/shared/isabella)



isabella  
CLUSTER

# Klaster Isabella



isabella  
CLUSTER



# Podnošenje poslova

# Opisivanje poslova

- ❖ Što se treba izvesti?
- ❖ Koji su argumenti?
- ❖ Gdje treba spremiti standardni izlaz i greške?
- ❖ Gdje se nalazi standardni ulaz?
- ❖ Koji podaci su potrebni za izvođenje poslova?

# Opisivanje poslova

- ❖ Gdje spremati privremene podatke i rezultate?
- ❖ Gdje je radni direktorij na čvorovima?
- ❖ Koje okoline varijable treba postaviti?
- ❖ Koliko procesora posao zahtjeva? Koju arhitekturu procesora?
- ❖ Koliko memorije, prostora na disku, vremena je potrebno?

# Ganglia

## ❖ Sustav za nadzor

- ♦ prikuplja informacije o čvorovima
- ♦ razmjenjuje informacije između frontenda i čvorova
- ♦ arhivira podatke

## ❖ Web sučelje

- ♦ prikaz grafova pojedinih parametara (CPU, memorija, mreža, disk)
- ♦ povijest: sat, dan, mjesec, godina

## ❖ <http://tannat.srce.hr/ganglia>





# Hvala na pažnji!

<http://isabella.srce.hr>

<http://wiki.cro-ngi.hr>