
Interactive European Grid

# PACX-MPI

**Sven Stork**
**HRLS, Stuttgart**

MpiGridUserCourse07, Dublin

# Content

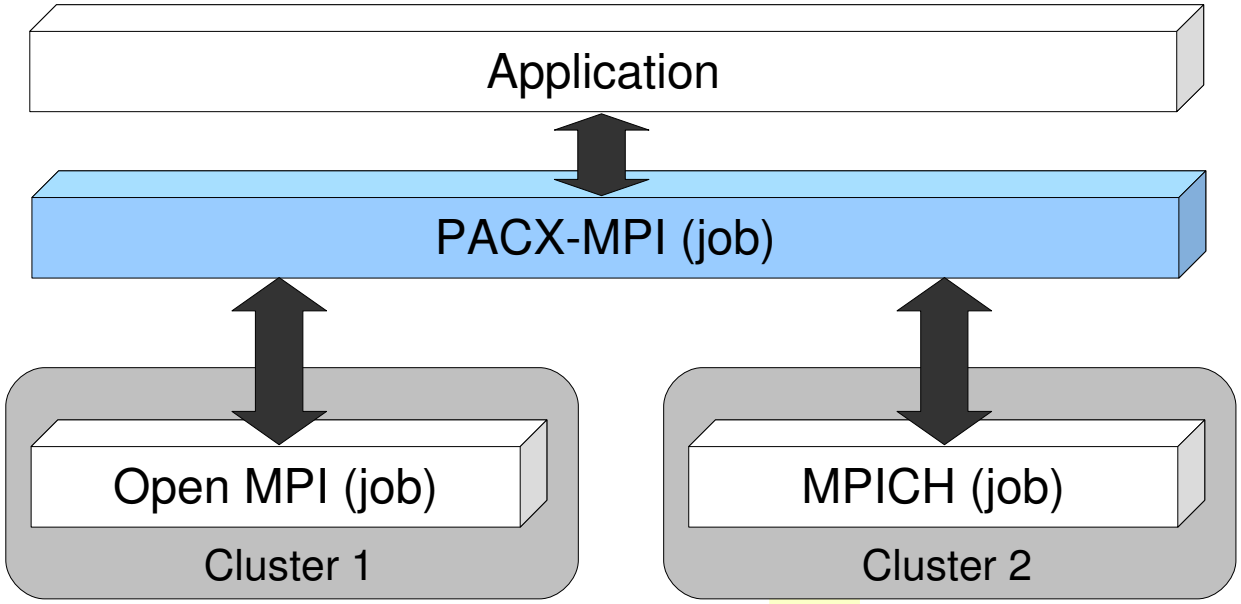
- PACX-MPI Overview
- PACX-MPI Examples

# PACX MPI – Overview

- A middleware for seamlessly run a MPI-application on a network of parallel computers
- originally developed in 1995 to connect Vector+MPP
- PACX-MPI is an optimized standard-conforming MPI- implementation, application just needs to be **recompiled(!)**
- PACX-MPI uses locally installed, optimized vendor implementations for cluster inter communication

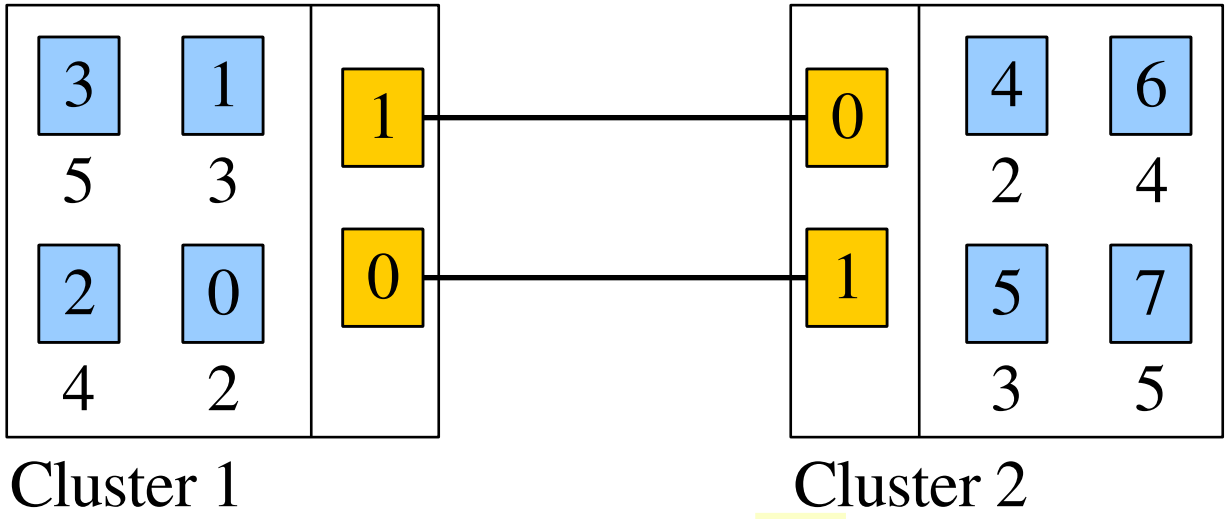
# PACX MPI – Design

- PACX-MPI start a MPI job in each cluster
- PACX-MPI “merge/manage” these MPI jobs internally and emulate transparently a bigger MPI job to the application



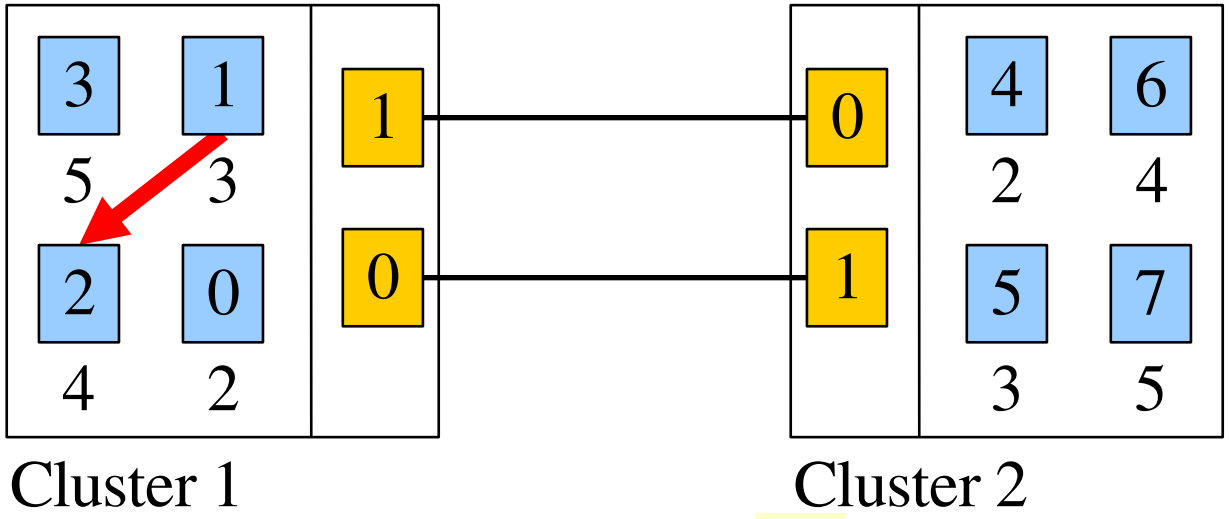
# PACX MPI – Design

- PACX-MPI maps the MPI process ranks of the big job to the processes of the clusters
- PACX-MPI start 2 additional,hidden MPI processes on the local MPI jobs for the external communication
- rank 0 of the local MPI jobs is always the **out_daemon**
- rank 1 of the local MPI jobs is always the **in_daemon**



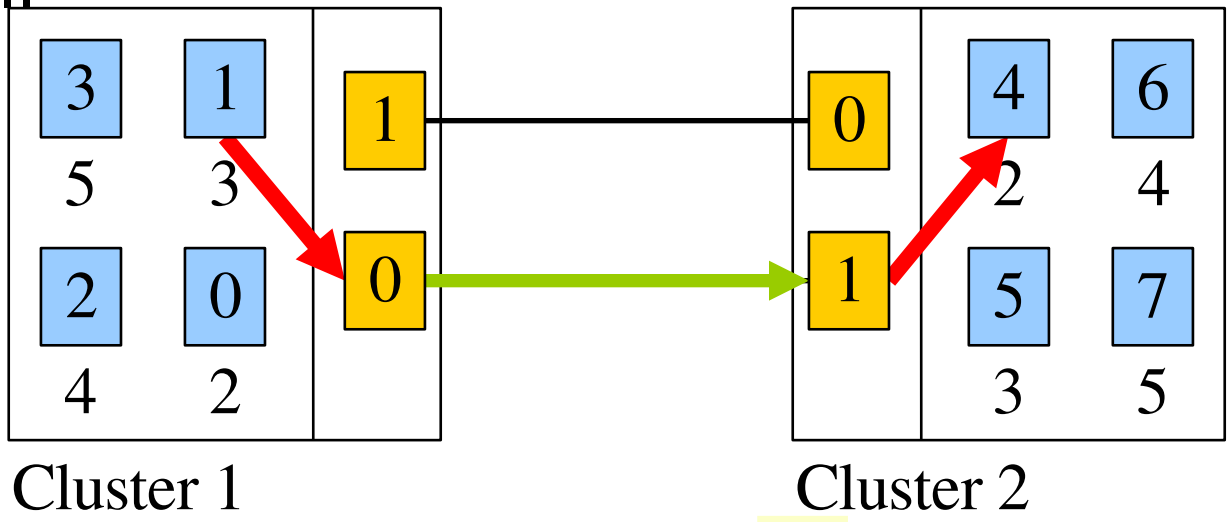
# PACX MPI – Design

- Internal Communication
  - communication between processes that resides in the same cluster is performed via the local, optimized MPI



# PACX MPI – Design

- External Communication
  - Send message to out_daemon using native MPI.
  - The **out_daemon** sends message to destination host over network using "a protocol"  (TCP, ATM, SSL)
  - The **in_daemon** sends message to destination with MPI.

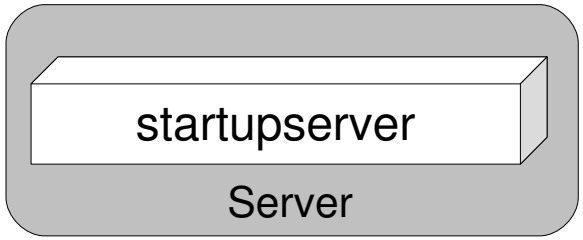


# PACX MPI – usage

- PACX-MPI compiler wrappers
  - pacxcc
  - pacxfc
  - ppacxcc
  - ppacxfc
- compiling with PACX
  - pacxcc -c hello.c
  - pacxcc -o hello hello.o

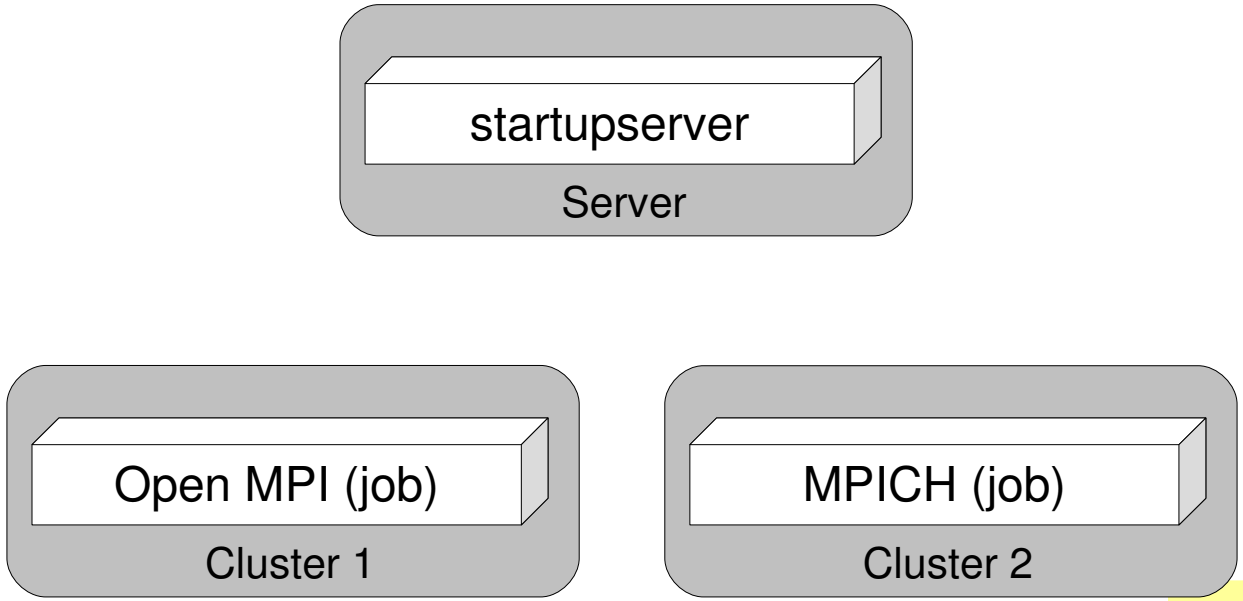
# PACX MPI – usage

- Running PACX MPI applications
  - Start the PACX ***startupserver*** with the number of clusters on a central reachable node
    - e.g. Server : ./startupserver 2



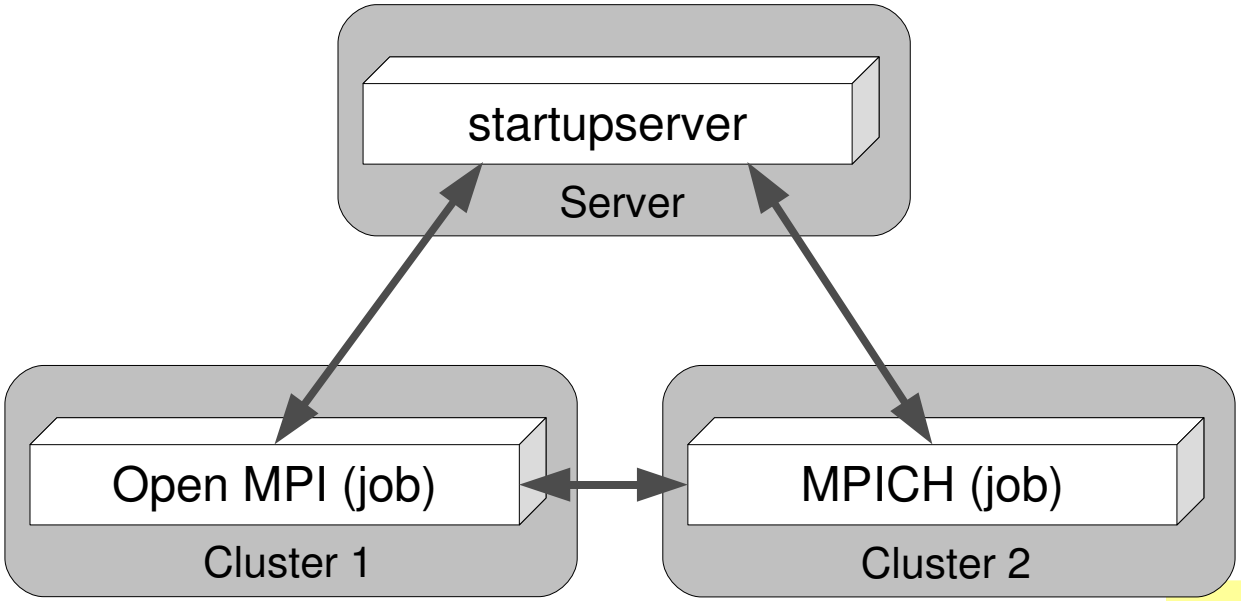
# PACX MPI – usage

- Running PACX MPI applications
  - start the application on each cluster with the required number of processes (+2 additional daemon processes)
    - e.g. Cluster 1 HN : mpirun -np 3 hello
    - e.g. Cluster 2 HN : mpirun -np 3 hello



# PACX MPI – usage

- Running PACX MPI applications
  - the applications connect and sync via the **startupserver** their corresponding connection information
  - the parallel programs (I/O daemons) connect to each other



# PACX MPI – Requirements + Configuration

- the ***startupserver*** requires requires one public reachable port, the port number can be specified as command line argument
- the first (n-1) jobs need 2 public reachable ports (at least the node where the in_daemon/out_daemon is running)
- the connection ports can be configure via the ".netfile" configuration file
- the requirements of the locally installed MPI implementations
- software installation
  - HN : PACX-MPI for the local MPI implementation (required for compilation)
  - CN : PACX-MPI is statically linked to the binary, therefore no extra software than the local MPI requirements need to be installed