



Hewlett Packard
Enterprise

SOFTWARE CHANGES TO ENABLE SLINGSHOT SUPPORT ON HPE SYSTEMS



Michael A. Raymond
MAY 2022

CONFIDENTIAL DISCLOSURE AGREEMENT

- The information contained in this presentation is proprietary to Hewlett Packard Enterprise and is offered in confidence, subject to the terms and conditions of a binding Confidential Disclosure Agreement (CDA)
- HPE requires customers and partners to have signed a CDA in order to view this training
- The information contained in this training is HPE confidential
- This presentation is NOT to be used as a 'leave behind' for customers and information may only be shared verbally with HPE external customers under NDA
- This presentation may be shared with Partners under NDA in hard-copy or electronic format for internal training purposes only
- Do not remove any classification labels, warnings or disclaimers on any slide or modify this presentation to change the classification level
- Do not remove this slide from the presentation
- HPE does not warrant or represent that it will introduce any product to which the information relates
- The information contained herein is subject to change without notice
- HPE makes no warranties regarding the accuracy of this information
- The only warranties for HPE products and services are set forth in the express warranty statements accompanying such products and services
- Nothing herein should be construed as constituting an additional warranty
- HPE shall not be liable for technical or editorial errors or omissions contained herein
- Strict adherence to the HPE Standards of Business Conduct regarding this classification level is critical

HPE CRAY MPI SOFTWARE ARCHITECTURE



HPE CASSINI FEATURES

- Hardware collectives offload
- Scalable start up
- Traffic Classes
- Security tokens
- Hardware resource control



COMMUNICATION LIBRARY CHANGES

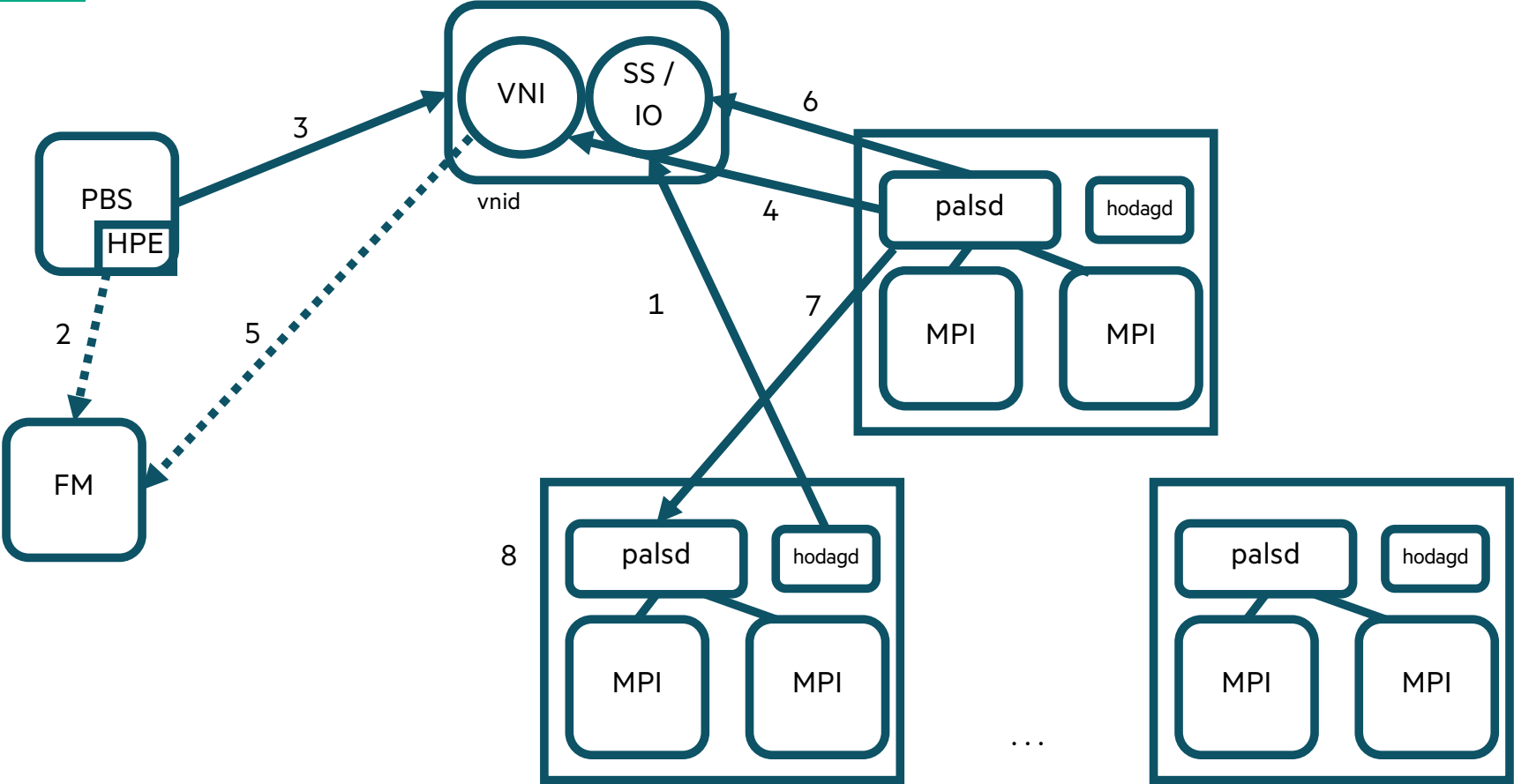
- SLINGSHOT_ environment variables pass settings from the launcher to the communication libraries
 - SLINGSHOT_VNIS
 - job-step and (optional) job VNIs
 - E.g. “42, 31415”
 - SLINGSHOT_SVC_IDS
 - In same order as _DEVICES
 - E.g. “3,2”
 - SLINGSHOT_DEVICES
 - Devices with services on them
 - E.g. “cxi0,cx1”, corresponds to -hsn0 & -hsn1 hostnames
- Slingshot features are available through libfabric
 - Traffic Classes will be available in version 2.0
 - Collectives will be in a future version

{

```
vni_key->vni = vni;
vni_key->svc_id = svc_id;
hints->ep_attr->auth_key = vni_key;
hints->ep_attr->auth_key_size =
    sizeof(*vni_key);
hints->addr_format = FI_ADDR_CXI;
mac_addr = local_macs[0];
sprintf(cxi_pid, "%d", lrank);
fi_getinfo(.., mac_addr, cxi_pid, FI_SOURCE,
    hints, &prov);
```

}

PBS / PALS CHANGES (PRESENT & FUTURE)



PALS COMMAND LINE OPTIONS

- Control Traffic Classes

\$ mpiexec -traffic-classes "dedicated_access,low_latency,bulk_data,best_effort" ...

- If the option is not used, then the app can use the TCs allowed by PBS
- If the option is used, then the intersection of the allowed and the mentioned TCs will be usable

- Control NIC resources

\$ mpiexec -network "def_txqs=32;max_txqs=256" ...

- The PALS configuration file has per-thread defaults and maximums for the resources
- When specified on the command line, the values will be shared among all the locally launched processes

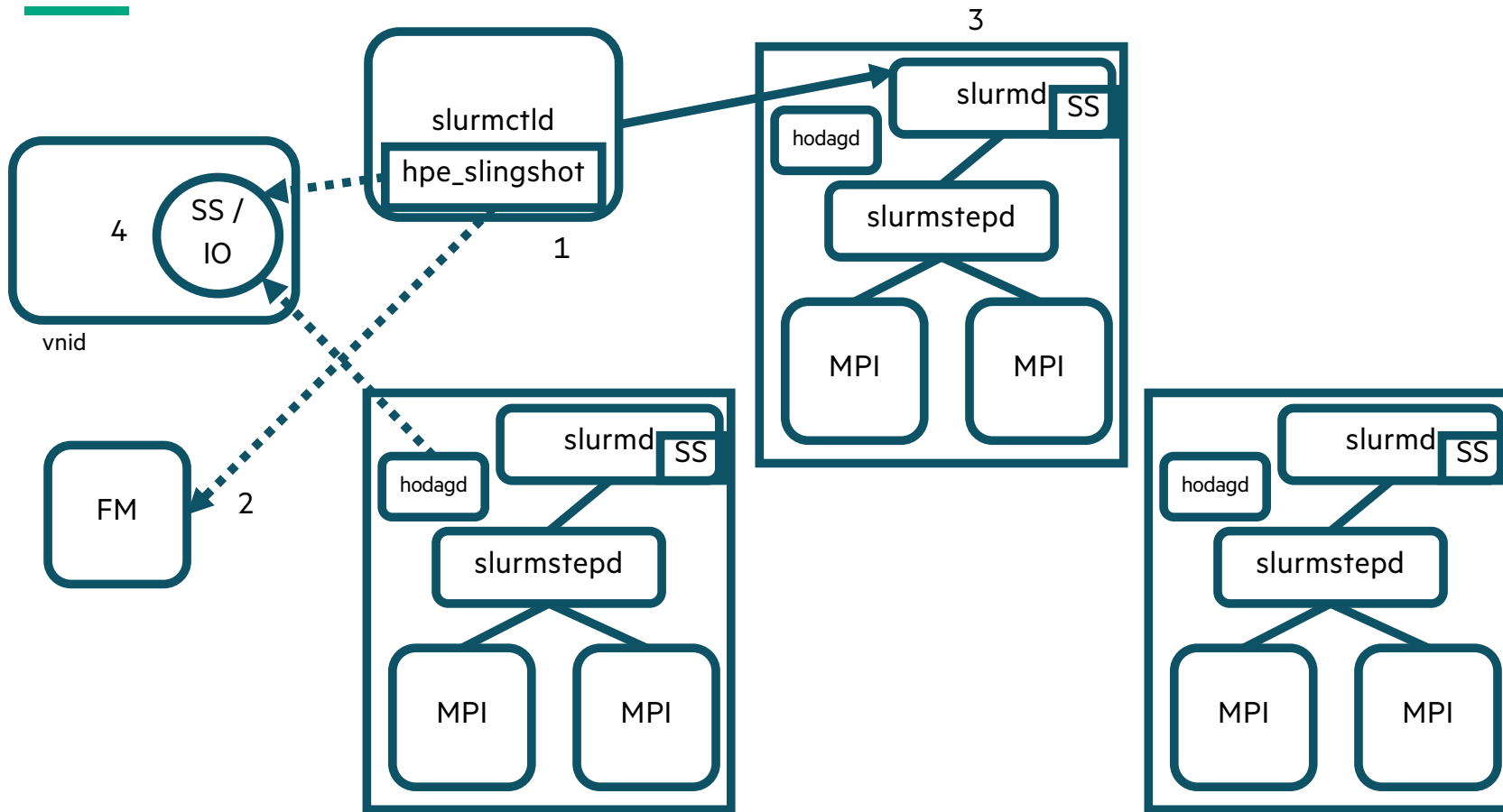
- Scalable Start Up / Instant On

\$ mpiexec -sstartup ...

- Future versions will support heterogenous PPNs and node configurations



SLURM CHANGES (PRESENT & FUTURE)



SLURM COMMAND LINE OPTIONS

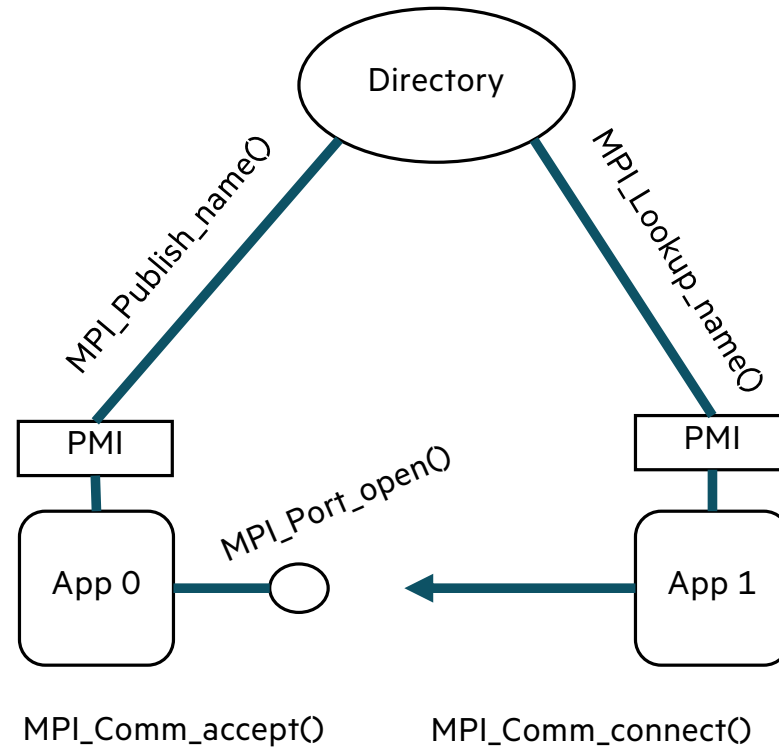
- Will add Traffic Class control in the future
- Control NIC Resources
 - `$ srun -network=res_txqs=64,res_eqs=32 ...`
 - The Slurm configuration file has per-thread defaults
 - The command line option sets the default per-thread allocations
- Will add Scalable Start up / Instant On control in the future



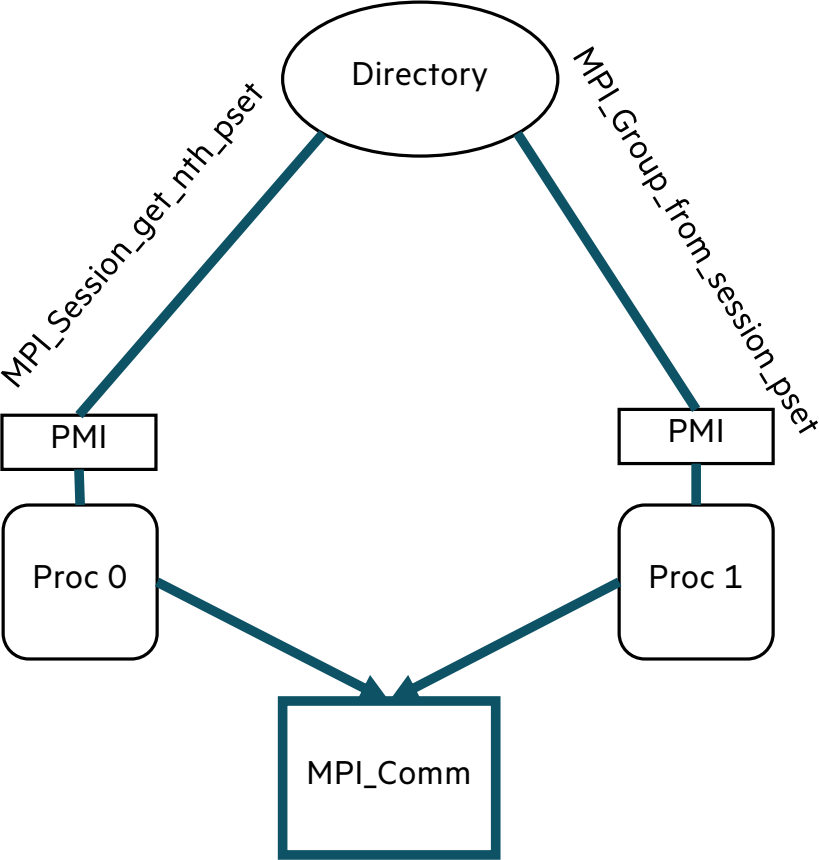
POTENTIAL NEW FEATURES



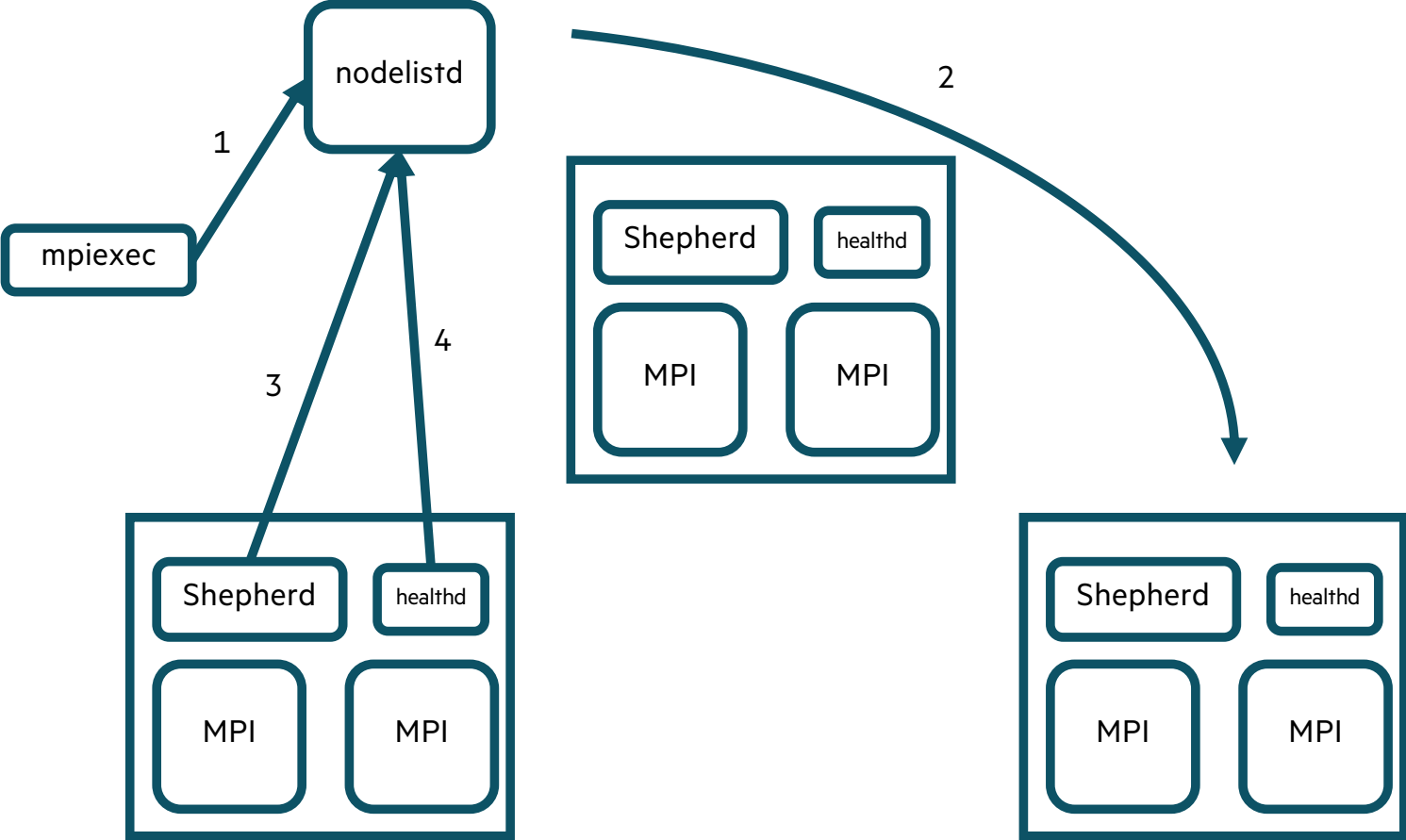
SUPPORT FOR MPI_PUBLISH_NAME / MPI_LOOKUP_NAME



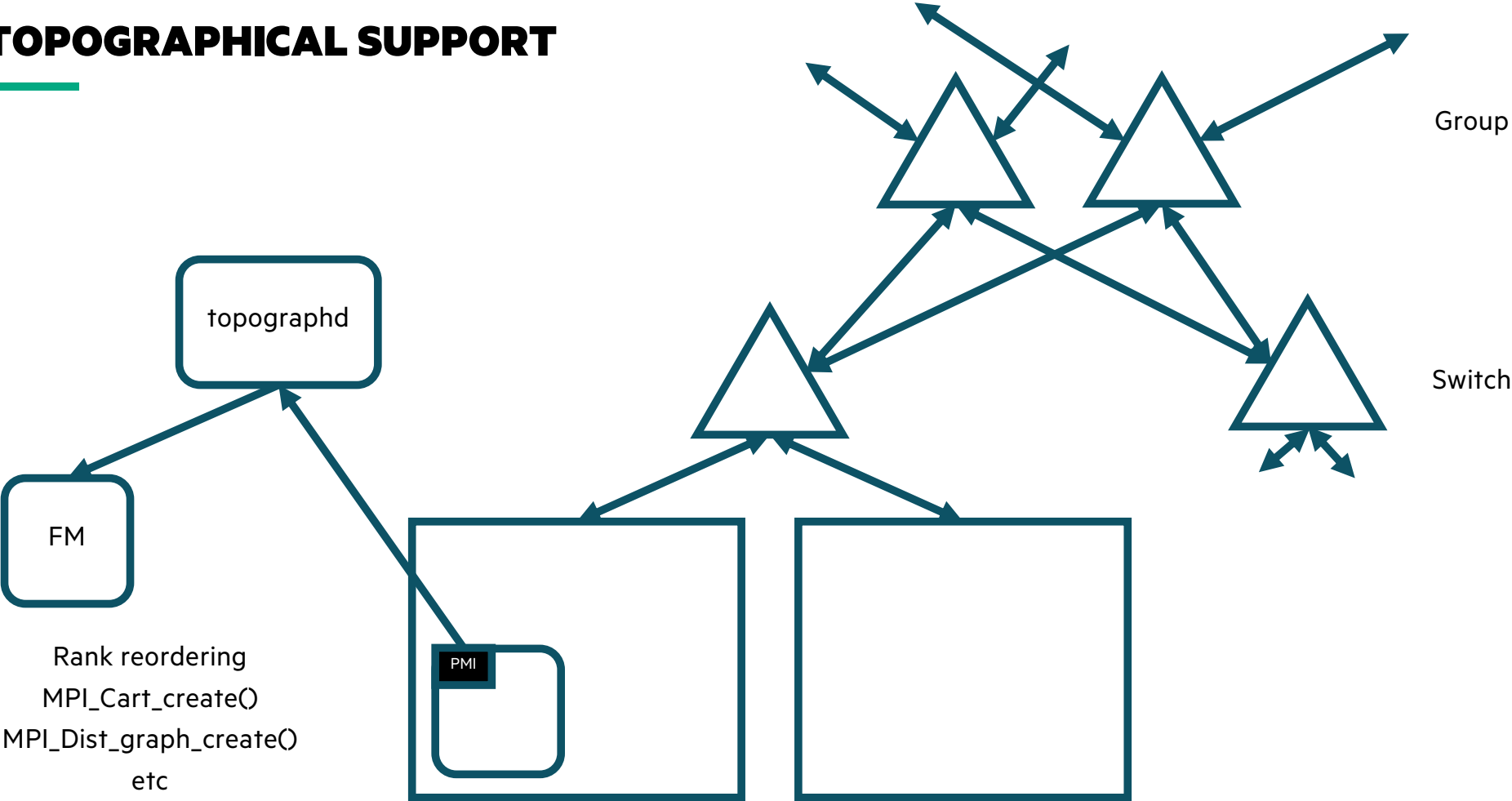
MPI 4.0 SESSIONS



PALS SUPPORT FOR NODE LISTS



TOPOGRAPHICAL SUPPORT



THANK YOU



<mraymond@hpe.com>



CONFIDENTIAL | AUTHORIZED