LA-UR-22-32131

Title:	PMIx and MPI Sessions, etc.
Author(s):	Pritchard, Howard Porter Jr.
Intended for:	Supercomputing - The International Conference for High Performance Computing, Networking, Storage and Analysis, 2022-11-13 (Dallas, Texas, United States)
Issued:	2022-11-17

Los Alamos
NATIONAL LABORATORY

Short Talk:
PMIx and MPI Sessions, etc.

Howard Pritchard (LANL)

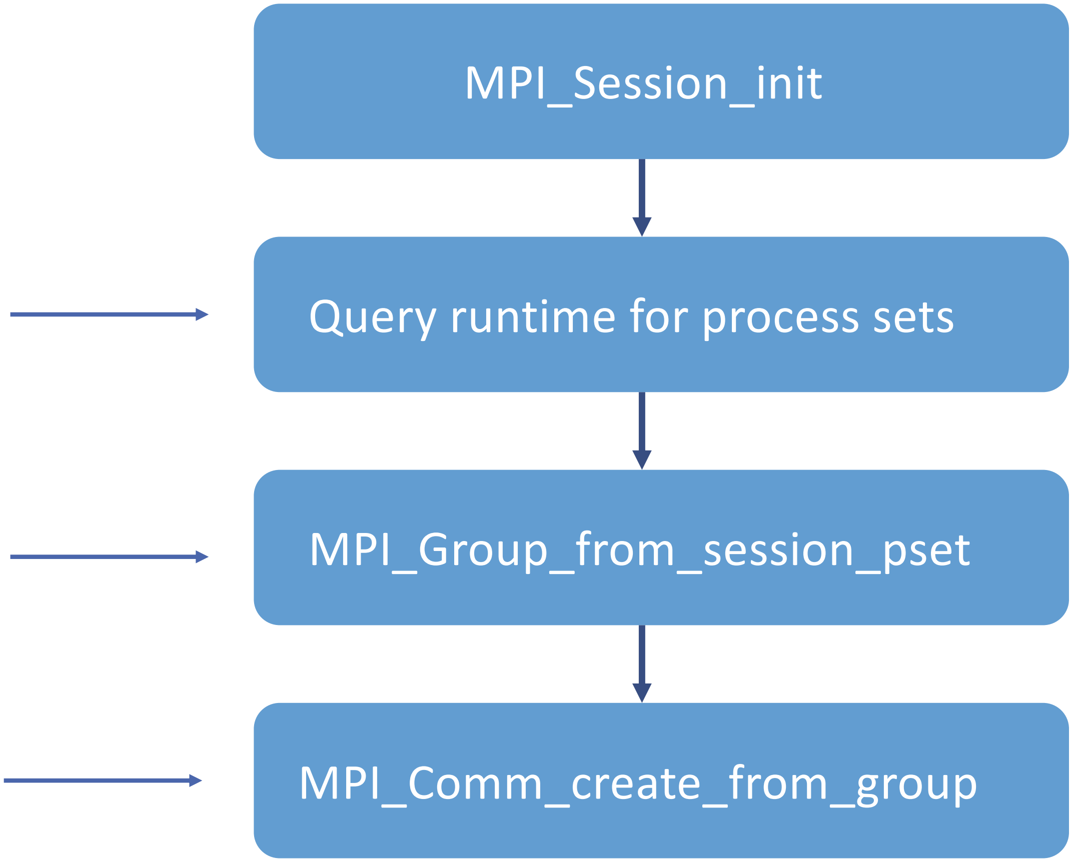
MPI Sessions (4.0 version) and PMIx

PMIx calls used currently (Open MPI)

PMIx_Query_info – PMIX_QUERY_NUM_PSETS,
PMIX_QUERY_PSET_NAMES

PMIx_Query_info - PMIX_QUERY_PSET_MEMBERSHIP
(not for mpi://world and mpi://self)

PMIx_Group_construct/destruct



MPI_Session_init

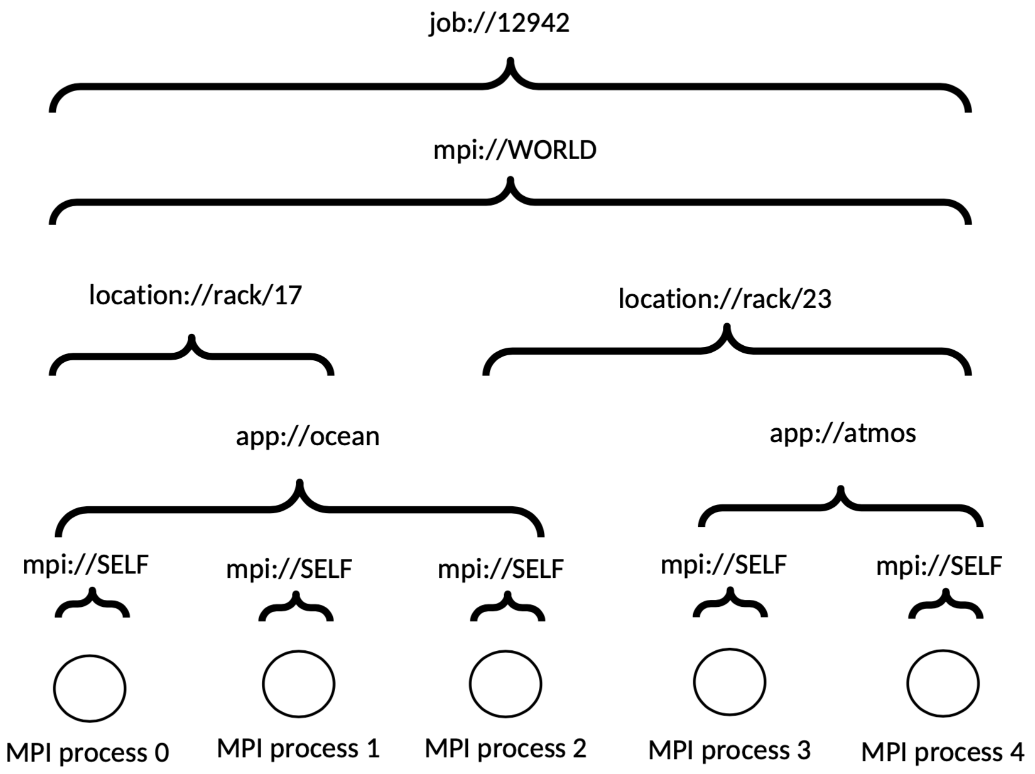
Query runtime for process sets

MPI_Group_from_session_pset

MPI_Comm_create_from_group

MPI Sessions (4.0 version) and PMIx Process Sets

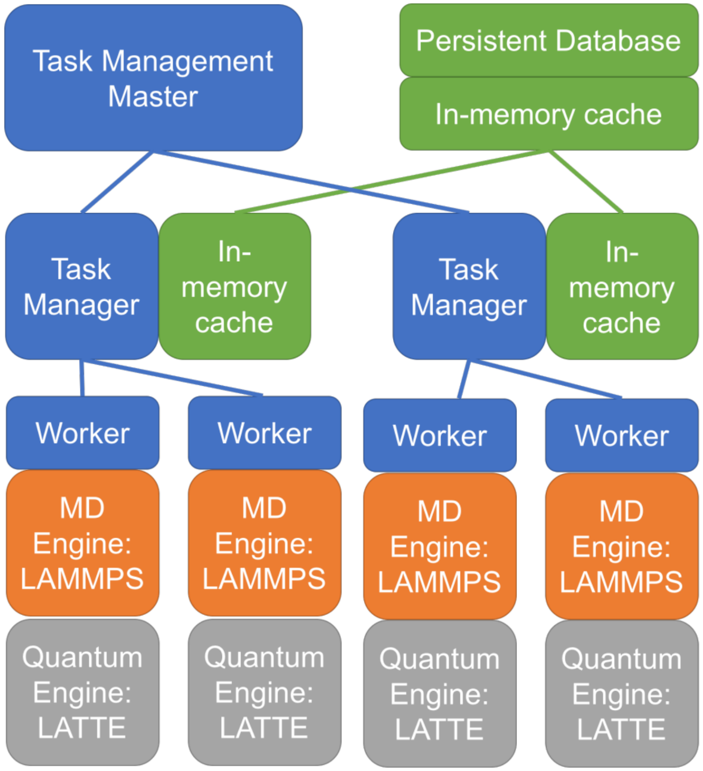
- Figure from the MPI 4.0 standard – illustrates possible process sets defined by the runtime at application launch
- This maps well to the PMIx Process Set definition (sec. 13.1 of PMIx 4.1 std)
- MPI Standard does have wording to indicate a runtime can create additional process sets after application launch
- *Upshot is PMIx Process Set definition does not map directly to the process set terminology in the MPI standard. This is okay.*



MPI 5 and Better Support for Malleable Applications

- Increasing number of HPC workflows could benefit from a more elastic runtime and resource scheduling environment

- MPI Sessions working group is exploring approaches within the context of MPI to expose capabilities of such a more elastic runtime to the application without introducing too much complexity



Exaalt infrastructure (ECP ADSE04)

PMIx Support for Malleable MPI Applications

PMIx 4 defines methods that, in principle, would provide much of the functionality needed to support approaches the MPI Sessions WG is considering for MPI5:

- Job management including resource allocation, job control, etc.
- Group management methods, e.g. *PMIx_Group_construct*, *PMIx_Group_invite*
  - PMIx groups can be mapped to MPI process sets
- Process creation – *PMIx_Spawn*

Numerous challenges, however, including PMIx server support for this functionality, limitations in resource management systems, etc.

What functionality would we want to expose through MPI verses more runtime oriented interfaces?

Some related Presentations at SC22

- Martin Schulz gave a talk - ***Adding Malleability to MPI: Opportunities and Challenges*** at the **ESPM2 2022** workshop on Monday
- A BOF **- Enabling I/O and Computation Malleability in High-Performance Computing -** which took place on Wednesday
- Some presentations at WORKS22 workshop on Monday