

BigHPC

A Management Framework for Consolidated Big Data and High-Performance Computing

Bruno Antunes, Júlio Costa (Wavecom); João Paulo, Ricardo Macedo, Mariana Miranda, Paula Rodrigues (INESC TEC); John Cazes, Stephen Harrell, Amit Ruhela (TACC & UT Austin); Samuel Bernardo, Miguel Viana (LIP)

Abstract

The BigHPC project aims at solving two main challenges that arise with the increased usage of Big Data Applications in HPC centers:

- Hardware and Software Heterogeneity
- Unsustained Storage I/O Performance

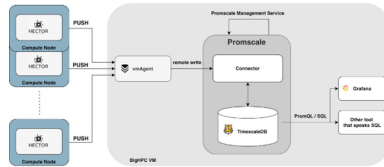
These problems are tackled by three main components:

- Monitoring Backend
- Virtualization Manager
- Storage Manager

Monitoring Backend

HECTOR is now capable of launching different processes

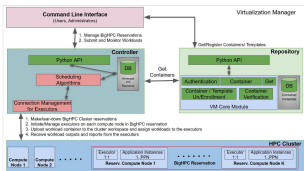
MONICA is the architecture that incorporates all services related with: Management, Visualization and Storage



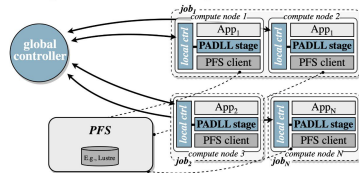
Virtualization Manager

Explored Designs Proposed Design Benefits

- | | |
|--|---|
| <ol style="list-style-type: none"> 1. Kubernetes 2. HTCondor 3. Jenkins 4. Indigenos | <ol style="list-style-type: none"> 1. Userspace Design 2. Generic (Job Schedulers, Sys Arch) 3. Lightweight and Performance 4. Support Persistence Services |
|--|---|



Storage Manager



PADLL stages mediate I/O requests from applications to the PFS

CHEFERD's global and local controllers tune stages holistically for sustained storage QoS

Integration

The previous three components are deployed following a GitOps approach. Using this framework, we are taking some advantages such as:

- Faster and more often deployment
- Easy and fast error recovery
- Infrastructure isolation

All interaction during application deployment does not require direct access to the internal infrastructure

Selected Publications

Protecting Metadata Servers From Harm Through Application-level I/O Control
 Rex-IO 2022 (co-located with IEEE Cluster)

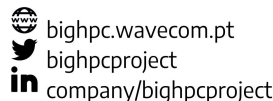
PAIO: General, Portable I/O Optimizations with Minor Application Modifications
 USENIX FAST 2022

Scaling Containerization on multi-Petaflops CPU and GPU HPC platforms
 HPBD 2021 (co-located with SRDS 2021)

Consortium



Follow



Funding

