Case Studies in Deploying Cluster Compatibility Mode

Tara Fly
Cray Management Services
Cray, Inc.

Introduction

Overview of Cluster Compatibility Mode

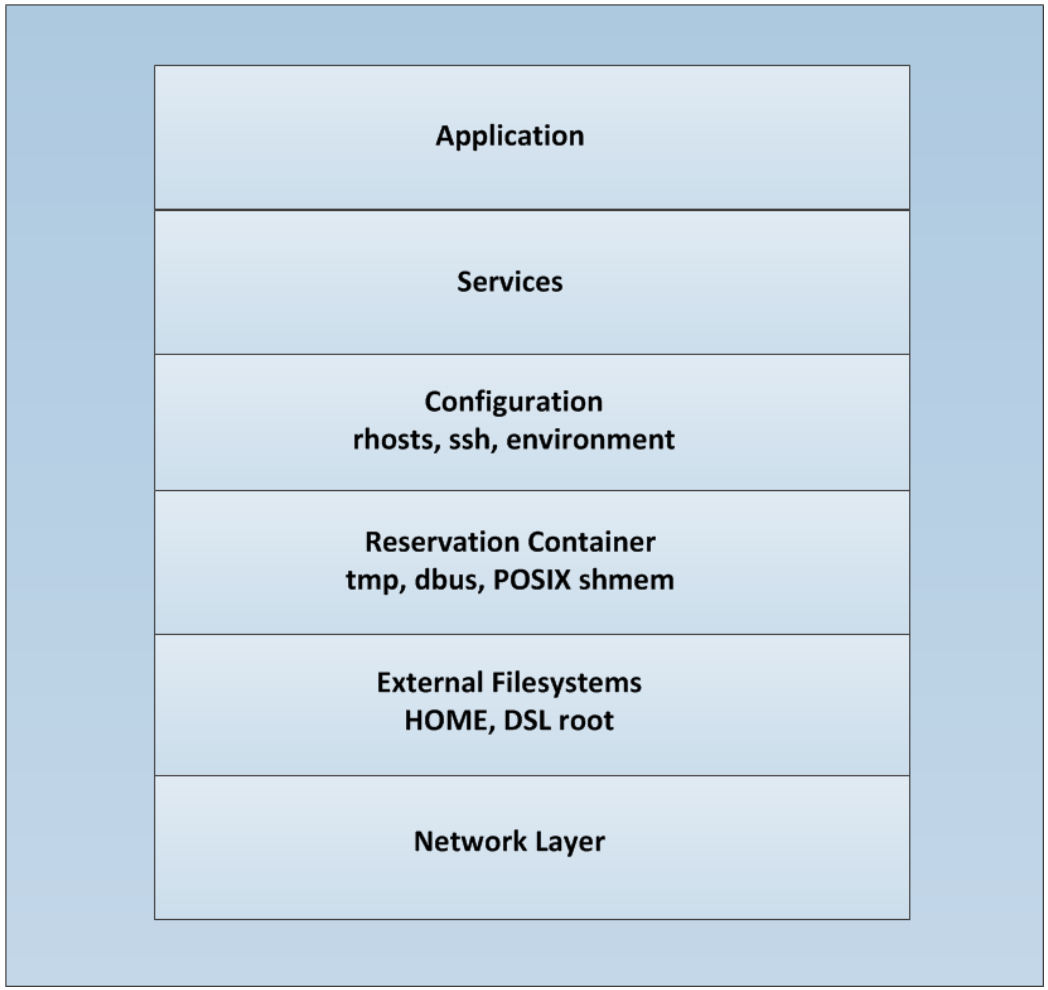
Deployment Scenarios

System Interactions

Cluster Compatibility Mode Design Goals

- Provide an environment to run ISV applications out-of-the-box without modification to the applications
- To ensure the capabilities necessary to run these ISV applications without sacrificing the performance of traditional Extreme Scalability Mode (ESM) applications
- To continue to optimize performance moving forward (ISV Application Acceleration)
- Scaling up to 2048 PEs per job instance

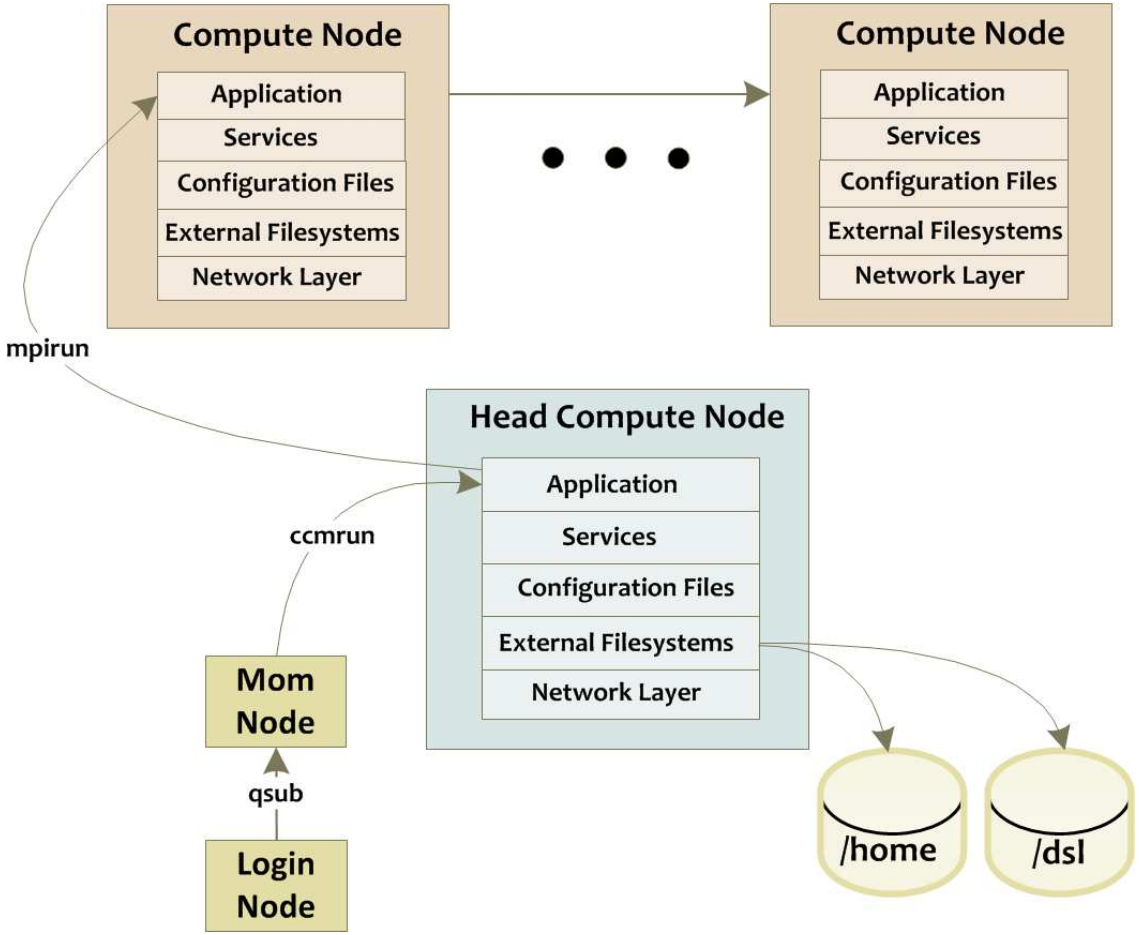
CCM Compute Node Environment



CCM User Interface

- **Primary interface through ccmrun**
  - Passes executable and any arguments along to compute node
  - Typical usage: ccmrun mpirun *application arguments*
  - Job launch in conjunction with workload manager
  - Users customize specifics of job, like any cluster
- **The ccmlogin command provided for X-authentication in interactive environment**

CCM Job Flow


Compute Node
Application
Services
Configuration Files
External Filesystems
Network Layer
Compute Node
Application
Services
Configuration Files
External Filesystems
Network Layer
mpirun
Head Compute Node
Application
Services
Configuration Files
External Filesystems
Network Layer
ccmrun
Mom Node
qsub
Login Node
/home
/dsl

Cluster vs. CLE Linux Differences

Cluster compute nodes run full Linux distribution
CLE compute node Linux image provided by shared root DSL

Cluster initializes Linux service on boot
DSL environment does not perform service initialization

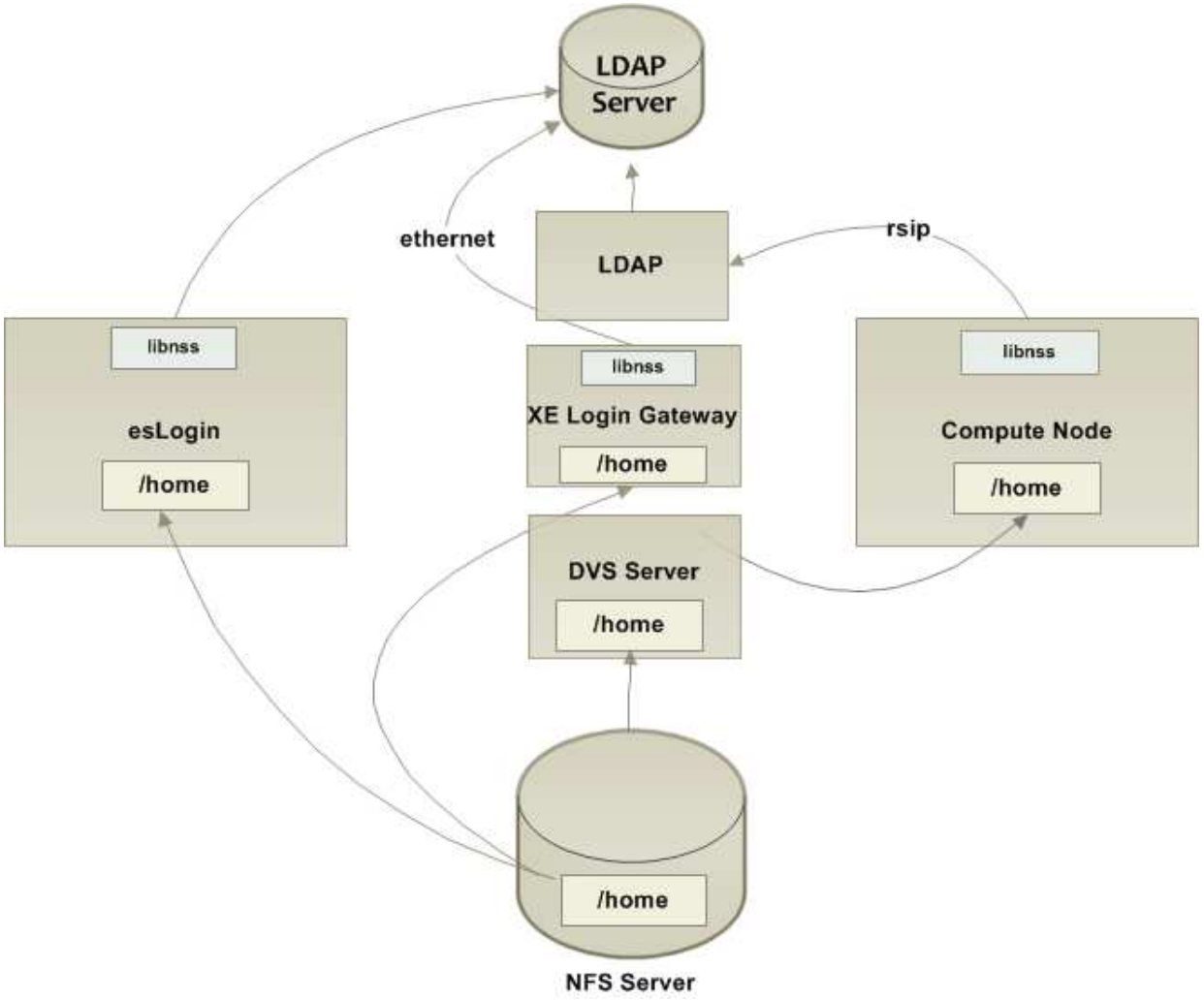
Workload managers perform job scheduling and launch
Workload managers perform scheduling; ALPS provides launch

Cluster node configuration implementation specific
CCM configuration management through shared root cnos class

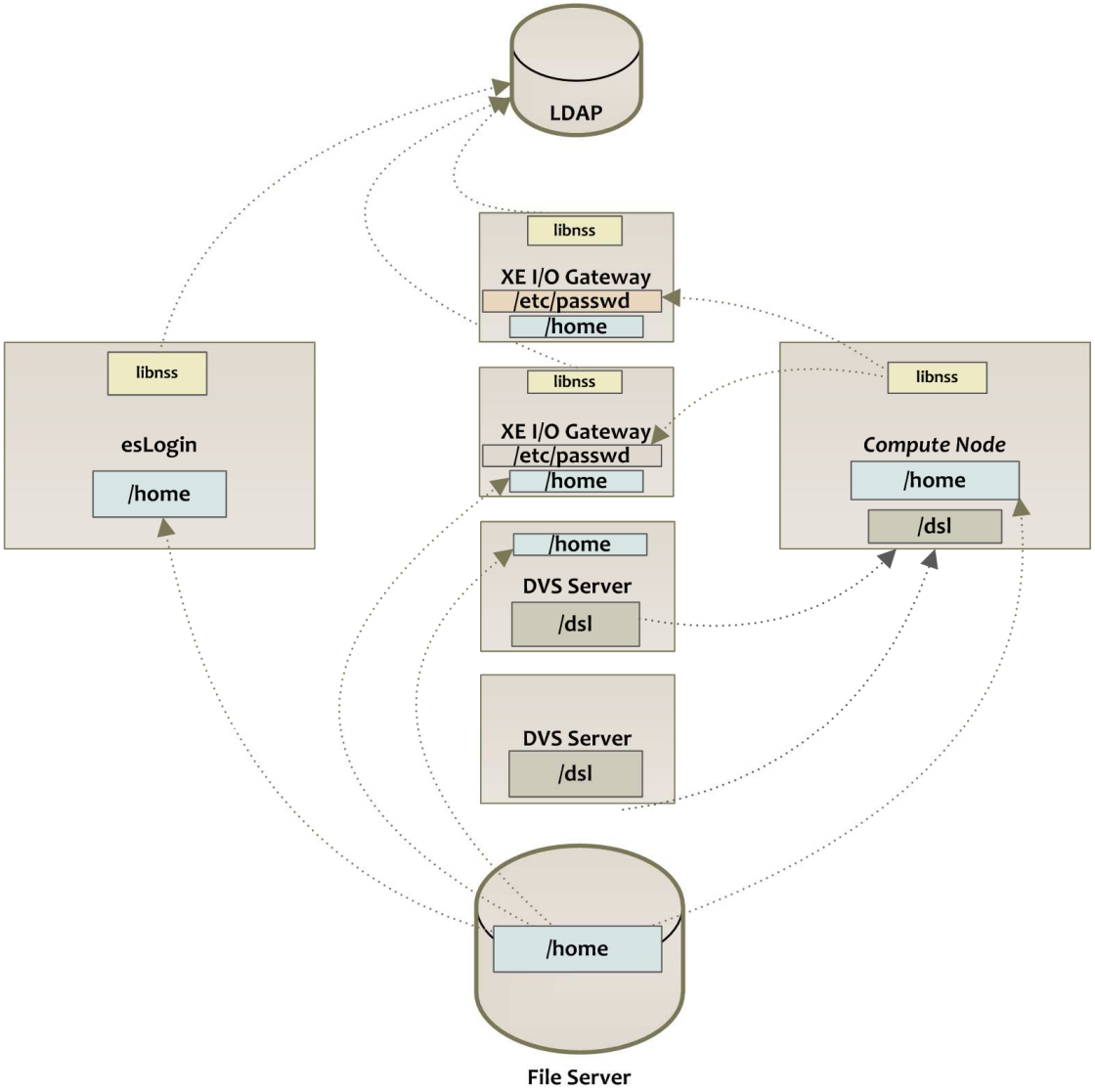
CCM Administration Overview

- CCM must run on top of DSL
- DSL is served by projecting shared root over DVS servers
- Configuration files managed by cnos class
- CNOS class is shared with all jobs targeting shared libraries
- Practical scaling constraints based on
  - Number of DVS servers in systems
  - Number of RSIP servers
  - Location of home directories
  - Name services authentication method chosen

CCM Configured with External LDAP



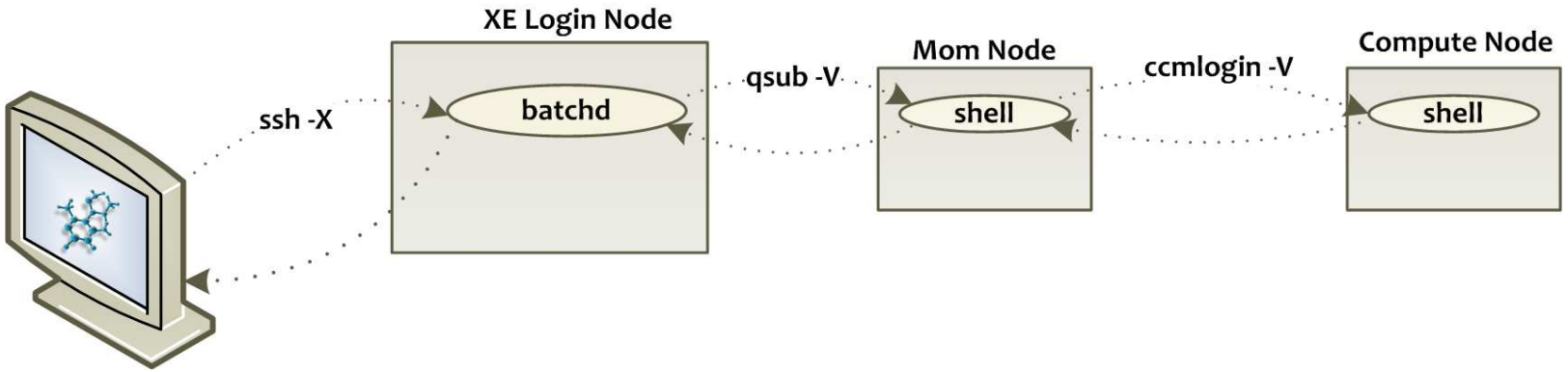
CCM Scaling Deployment



CCM Administration Best Practices

- **Use of CCM local mounts to isolate LDAP to just CCM jobs**
  - nsswitch.conf mounted only for CCM jobs
  - provide static password users to compute nodes
- **NIS slave server must be configured on HSN**
- **Provide /home in a scalable location**
- **Use Lustre storage for working directories of apps**
- **SSH provides greater security**
- **CCM will not guarantee xinetd sessions are isolated to a given user, although it will prevent accidental spillage**
- **Root squash home environment requires at least world-execute on home directory for password-less RSH configuration**
  - Users can alternately set up their own .rhosts file
- **CSA accounting configuration needs to be done carefully or could break CCM ISV Application Acceleration**

Graphical User Interface Support



Considerations for ccmlogin

- **Some site document methods including:**

  env > ~/.ssh/environment
  ccmlogin

- **This can cause application conflicts when running more than one application at a time**
- **Breaks X-forwarding**
- **Users own removing their environment file or may have unexpected future results**
- **ccmlogin –V is slightly better. Still single user restricted.**
- **Best method is through –x and –prefix options with MPI**

ISV Applications Tested Under CCM

- Abaqus Explicit
- Abaqus Implicit
- ANSA
- ANSYS CFX
- ANSYS FLUENT
- ANSYS Mechanical
- CFD++
- CHARMm
- EnSight
- AVL/FIRE
- GASP
- Gaussian
- iSight
- LSDYNA
- MATLAB
- MSC/NASTRAN
- MSM Amorphous Cell
- MSM Castep
- MSM Discoverer
- MSM Dmol3
- PowerFLOW
- Radioss
- Star-CD
- Star CCM+

Performance Numbers

XE6, native	XE6, Platform MPI, CCM w/ IAA	IB Cluster, QDR	IB Cluster, DDR
1.43	2.57	1.73	1.76

IMB PingPong latency, 8 byte messages

All times in microseconds

Bandwidth Comparison