Integrating hooks and tools for better manageability of HPC Cluster

By Munira Hussain and Calvin Jacob

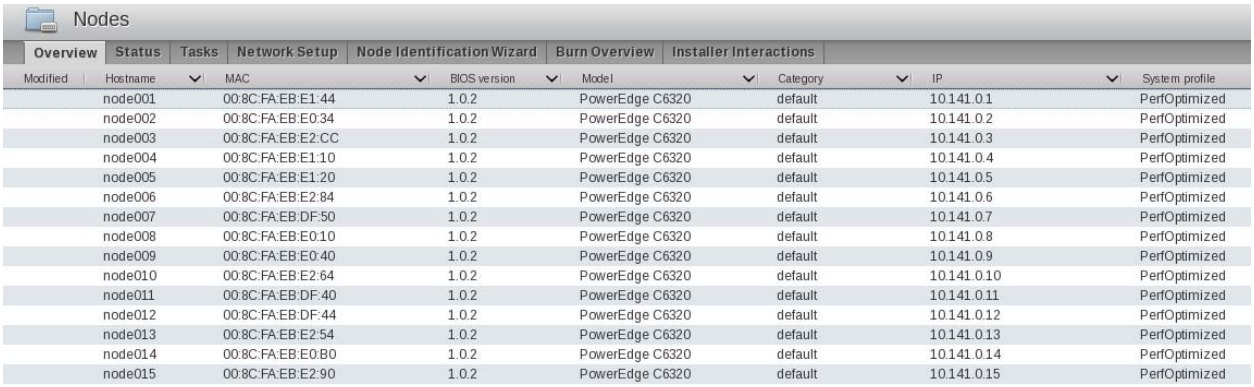
Managing local and remote server nodes in a cluster at scales of tens to thousands of nodes is always a challenge. In order to reduce the cluster-management overhead and simplify setup of cluster of nodes, admins seek the convenience of a one-snapshot glimpse. Rapid changes in technology makes ease of management, tuning, customization, changing and updating settings an ongoing reality as infrastructure is refactored and refreshed.

To simplify some of these challenges, it is important that hardware management be fully integrated into the solution for cluster management. . The integration between Bright Cluster Manager 7.1 and Dell PowerEdge Haswell based servers depict a leading example of what is achievable today. Critical to this integration is Integrated Dell Remote Access Controller (IDRAC). IDRAC is embedded into the motherboards of Dell servers for in-band and out-of-band system management and can display and modify BIOS settings as well as perform firmware updates through the Life Cycle Controller and remote-console. Collectively, each server's in-depth system profile information is gathered using system tools and utilities and is available in a single graphical user interface for the ease of administration, thus reducing the need to physically access servers directly.

Figure *1*. BIOS-level integration between Dell PowerEdge servers and Bright Cluster Manager 7.1.

Figure 1 (above) depicts the configuration setup for a single node in the cluster. The fabric can be accessed via the dedicated iDRAC port or shared with the LAN-on-Motherboard capability. The cluster administration fabric is configured at the time of deployment using Bright Cluster Manager 7.1. The system profile of the server is captured in an XML-based schema file that gets imported from the iDRAC using the racadm commands. Thus relevant data (e.g., optimal system BIOS settings, boot order, console redirection, network configuration, etc.) is parsed and displayed on the cluster dashboard of the graphical user interface. By reversing this process, it is possible to apply BIOS settings onto servers Drop-down choices for setting the BIOS level system profiles are available from the Bright GUI for selection. These choices are then stored in an updated XML-based schema file on the head node, and pushed out to the appropriate nodes during reboots.

Figure *2*. Snapshot of the Cluster Node Configuration via Bright Cluster Manager 7.1.



Modified	Hostname	MAC	BIOS version	Model	Category	IP	System profile
	node001	00:8C:FA:EB:E1:44	1.0.2	PowerEdge C6320	default	10.141.0.1	PerfOptimized
	node002	00:8C:FA:EB:E0:34	1.0.2	PowerEdge C6320	default	10.141.0.2	PerfOptimized
	node003	00:8C:FA:EB:E2:CC	1.0.2	PowerEdge C6320	default	10.141.0.3	PerfOptimized
	node004	00:8C:FA:EB:E1:10	1.0.2	PowerEdge C6320	default	10.141.0.4	PerfOptimized
	node005	00:8C:FA:EB:E1:20	1.0.2	PowerEdge C6320	default	10.141.0.5	PerfOptimized
	node006	00:8C:FA:EB:E2:84	1.0.2	PowerEdge C6320	default	10.141.0.6	PerfOptimized
	node007	00:8C:FA:EB:DF:50	1.0.2	PowerEdge C6320	default	10.141.0.7	PerfOptimized
	node008	00:8C:FA:EB:E0:10	1.0.2	PowerEdge C6320	default	10.141.0.8	PerfOptimized
	node009	00:8C:FA:EB:E0:40	1.0.2	PowerEdge C6320	default	10.141.0.9	PerfOptimized
	node010	00:8C:FA:EB:E2:64	1.0.2	PowerEdge C6320	default	10.141.0.10	PerfOptimized
	node011	00:8C:FA:EB:DF:40	1.0.2	PowerEdge C6320	default	10.141.0.11	PerfOptimized
	node012	00:8C:FA:EB:DF:44	1.0.2	PowerEdge C6320	default	10.141.0.12	PerfOptimized
	node013	00:8C:FA:EB:E2:54	1.0.2	PowerEdge C6320	default	10.141.0.13	PerfOptimized
	node014	00:8C:FA:EB:E0:B0	1.0.2	PowerEdge C6320	default	10.141.0.14	PerfOptimized
	node015	00:8C:FA:EB:E2:90	1.0.2	PowerEdge C6320	default	10.141.0.15	PerfOptimized

Figure 2 is a screenshot from Bright Cluster Manager 7.1 that shows BIOS version and system profile information for a number of Dell PowerEdge C6320 servers. This is a particularly useful overview as inappropriate settings and versions can be easily and rapidly identified.

An example usage of this would be when new servers are added or replaced in a cluster. The above integration will help to determine that all servers have similar homogenous performance as this can be determined by confirming settings of BIOS versions and firmwares, system profile and other BIOS tuning configurations.

Similarly this is also useful for customers that may need specific settings applied on their servers that may not be a default setting. For example codes that are latency sensitive may require custom profile with C-States disabled. These servers can be categorized into a node group and specific BIOS parameters can be applied for that set of group.

This tightly coupled BIOS level integration, between Bright Cluster Manager 7.1 and Dell PowerEdge Haswell based servers, delivers capabilities that provide a significantly enhanced solution offering for Dell customers. As a validated and tested solutions on Dell hardware, they provide seamless operation and administration of Dell HPC clusters.

References:

1. http://www.brightcomputing.com/Bright-Cluster-Manager
2. http://en.community.dell.com/techcenter/systems-management/w/wiki/3204.dell-remote-access-controller-drac-idrac
3. http://www.brightcomputing.com/Linux-Cluster-Architecture
4. http://en.community.dell.com/techcenter/high-performance-computing/b/general_hpc/archive/2014/09/23/bios-tuning-for-hpc-on-13th-generation-haswell-servers
5. http://en.community.dell.com/techcenter/high-performance-computing/b/general_hpc/archive/2015/04/29/linpack-benchmarking-on-a-4-nodes-cluster-with-intel-xeon-phi-7120p-coprocessors