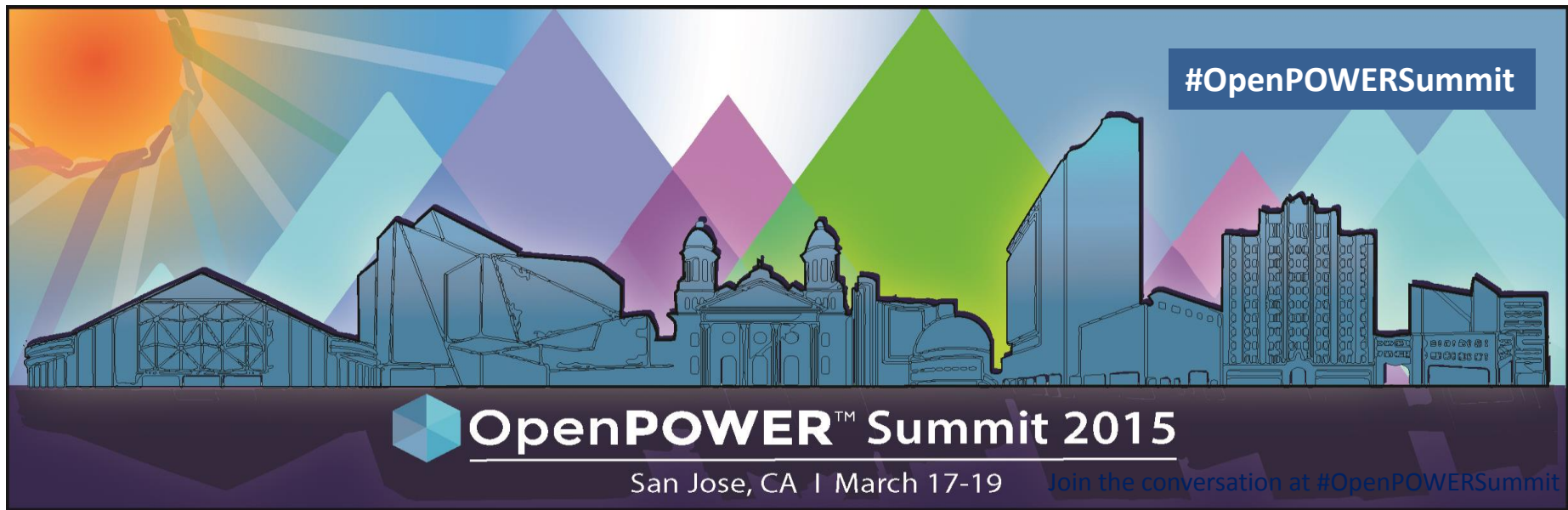




IBM Platform Computing : infrastructure management for HPC solutions on OpenPOWER

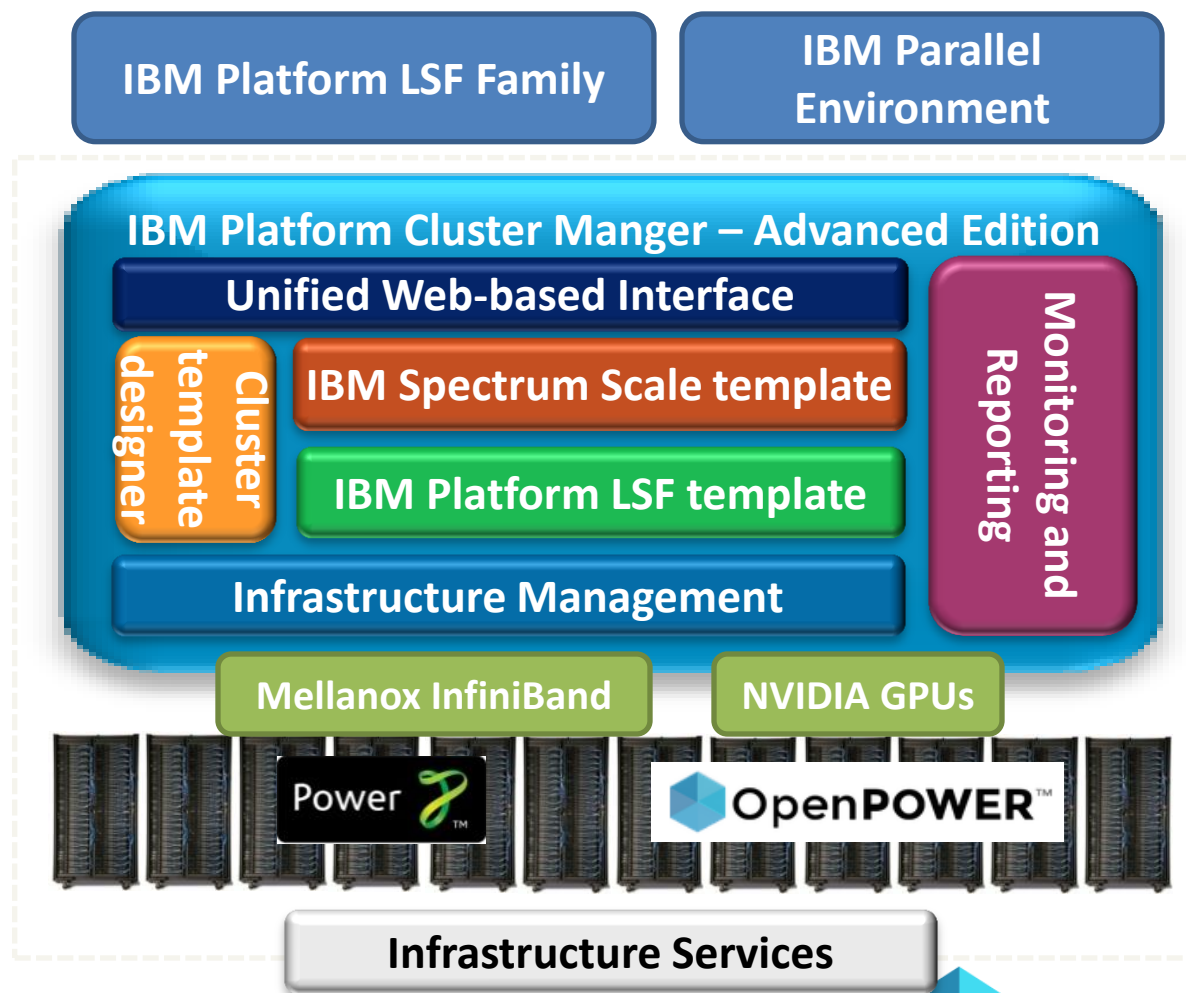
Jing Li, Software Development Manager
IBM



Scale-out and Cloud Infrastructure Management Needs

| | |
|--------------------------------------|--|
| Traditional HPC cluster | <ul style="list-style-type: none"> • Provisioning & monitoring system for scale-out computing infrastructures |
| Multi-tenants HPC environment | <ul style="list-style-type: none"> • Technical computing capacity to multiple departments, projects, and users (multi-tenants) • Self service capability |
| Data Management | <ul style="list-style-type: none"> • Elastic Storage Management and monitoring • Elastic Storage Appliance management and monitoring |
| Cloud Infrastructure | <ul style="list-style-type: none"> • VM, virtual network, & virtual storage management • Underlying infrastructure (undercloud) management • Configuration management |
| Capacity overflow | <ul style="list-style-type: none"> • Automated cloud bursting capabilities • Flexibility and benefits of VMs |
| Big data clusters | <ul style="list-style-type: none"> • Multi-tenancy • Enterprise class system management capabilities (audit, Role Based Access Control etc.) |

Architecture Overview



IBM Platform Cluster Manager

Overview

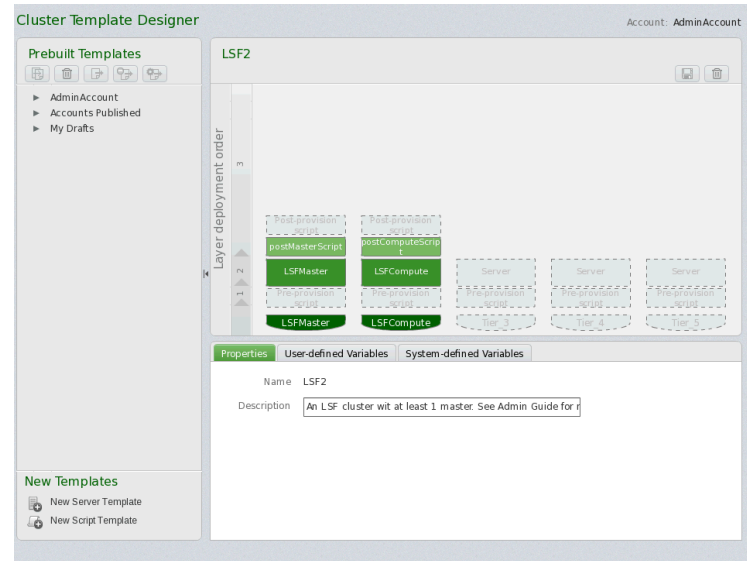
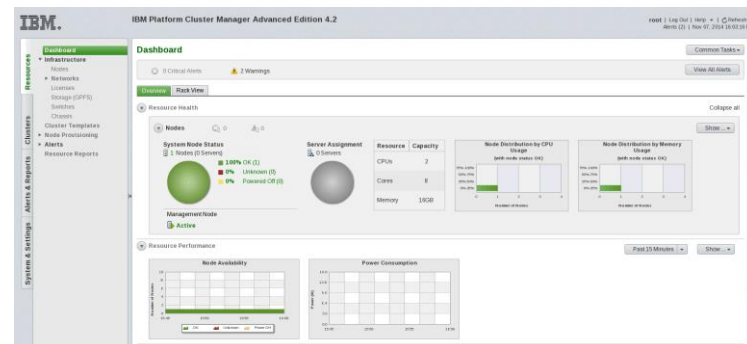
Powerful lifecycle management for scale-out cluster environments

Key Capabilities

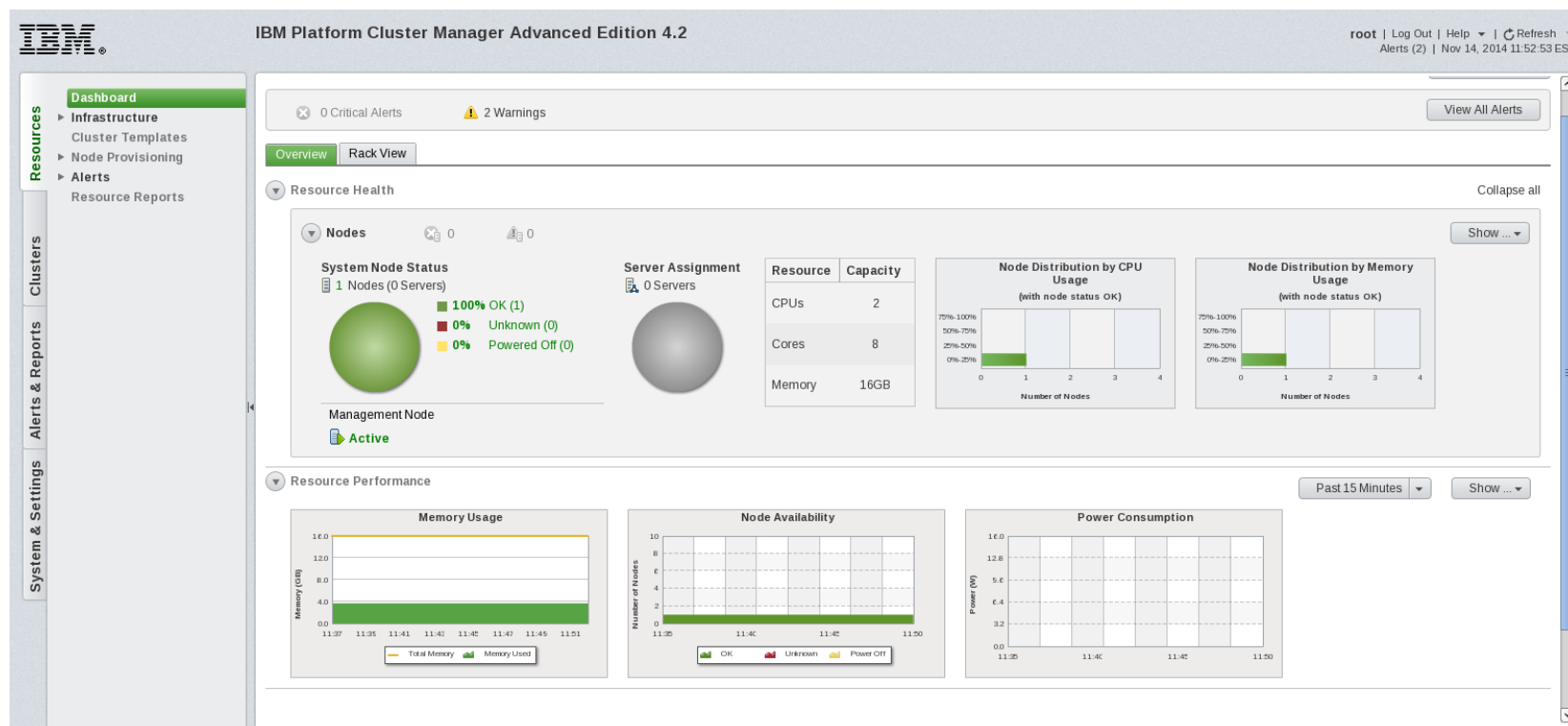
- Simplified management with cluster template designer
- Scales from single clusters to complex multi-team environments
- Robust, scalable alerting and reporting
- Automated infrastructure management – one-click cluster deployment
- Spectrum Scale cluster support

Benefits

- Faster time to cluster readiness
- Unified interface for management and monitoring
- Increased administrator productivity
- Single infrastructure supporting multiple business needs



Infrastructure at a glance



- Single interface for management and monitoring of multiple clusters
- Dashboard provides overview of resources, allocations

- 2D representation of the data center (racks, nodes)
- It allows administrator to quickly examine the status of individual nodes
- Admin can drill into node details by clicking the node
- Chassis console can be launched from the rack view
- Overview of entire cluster at a glance





OpenPOWER™

Managing node personalities

Jobs

Resources

System & Settings

- Dashboard
- ▼ **Devices**
 - Nodes
 - Chassis
 - Switches
- Unmanaged Devices
- Licenses
- ▼ **Node Provisioning**
 - Provisioning Templates**
 - Image Profiles
 - Network Profiles
 - Networks
 - OS Distributions
 - Kit Library
- ▼ **Resource**
 - Alert D
 - Trigger
- Application

Modify Provisioning Template

Name: RHEL6.2-stateful-compute

Description:

Node Name Format: comp#NNN

Image Profile: rhels6.2-x86_64-stateful-compute

OS: rhels6.2-x86_64

Provision Method: Stateful Package-based

Description: The rhels 6.2 x86_64 stateful compute image profile

Network Profile: default_cn

| Interface | Network | Subnet | IP Range |
|-----------|-----------|---------------|-------------------------------|
| eth0 | provision | 192.168.152.0 | 192.168.152.3-192.168.152.254 |

Hardware Profile: default_flex_x

Type: ipmi

Provisioning templates, image and network profiles can be easily managed all through the GUI.

Modify Network Profile

Name: default_cn

Description:

Network Interfaces
Add network interfaces to the network profile. You must add at least one network interface.

| Interface | Network | Subnet | IP Range | Type | Configuration Command |
|-----------|-----------|--------------|--------------|----------|-----------------------|
| eth0 | provision | 192.168.1... | 192.168.1... | Ethernet | - |

Primary Interface: eth0

Installation Interface: ☒ Same as Primary Interface ☐ Other

Modify Image Profile

Image Profile: rhels6.2-x86_64-stateful-compute

General

Name: rhels6.2-x86_64-stateful-compute

Description: The rhels 6.2 x86_64 stateful compute image profile

Provision Method: Stateful Package-based

Node Type: Compute Node

Built-in: No

OS Distribution: rhels6.2-x86_64

OS Update: <None>

Disk Partition: ☒ Use OS Default ☐ Use Customized Script

Boot Parameters:

Hardware management and monitoring

- Node detail is monitored with system & performance
- Management capabilities include: power cycling, firmware updates, OS reboot, reprovision, synchronize, node LED control, BMC, SSH, VNC console
- Integrated network switch and chassis monitoring
- Integrated Spectrum Scale monitoring

The screenshot displays the IBM Platform HPC 4.2 web interface. On the left, a VNC console window shows a terminal session with commands like 'firefox', 'ls', and 'cat /etc/passwd'. The main interface features a sidebar with navigation options: Dashboard, Infrastructure (Nodes, Networks, IP Pools, Reserved IPs, Licenses, Switches), Jobs, and Templates. The 'Nodes: List' table shows the following data:

| Name | Power Status | Workload Agent | Total Job Slots | Job Slots In Use | Provision Stat... | Synchronizati... | Node Type | CPU Usage | Free Memory ... | Node Groups | Tags |
|---------------|--------------|----------------|-----------------|------------------|-------------------|------------------|-----------------|-----------|-----------------|-------------|-------|
| bay2node | on | OK | 8 | 0 | defined | syncd | Compute Node | 0% | 47744 | - | - |
| bay3node | on | OK | 16 | 0 | failed | out-of-sync | Compute Node | 0% | 64000 | - | Ta... |
| bay4node-test | unknown | Unavailable | 1 | 0 | defined | failed | Compute Node | - | - | - | Ta... |
| pok007 | - | OK | 8 | 0 | provisioned | syncd | Management N... | 1% | 44575 | - | - |

Below the table, the 'Node Name: bay2node' section provides detailed monitoring for this specific node. It includes tabs for Summary, Performance, Jobs, Alerts & Events, and Properties. The Performance tab is active, showing several charts: CPU Usage (0-100%), Memory Usage (0-500 GB), Available Swap Space (0-300 GB), Login Users (0-50), Idle Time (0-2.4K), and Load Levels (0-1.0). The charts show data for the last 15 minutes.

Alerts highlight potential issues in the cluster

- Fully customizable; alerts can be defined using any monitored metrics
- Alert can trigger an automated pre-defined action
- Alert history shows the detail of the triggered alert

The screenshot shows the IBM Platform HPC 4.2 web interface. The left sidebar contains navigation menus for Dashboard, Infrastructure, Networks, IP Pools, Reserved IPs, Licenses, Switches, Chassis, Node Provisioning, Provisioning Templates, Image Profiles, Network Profiles, OS Distributions, Kit Library, Application Templates, Alerts, Alert Policies, Alert History, and Resource Reports. The main content area displays the 'Alert Policies' section. At the top, there are buttons for 'New', 'Delete', 'Modify', 'Enable', and 'Disable'. Below this is a table listing various alert policies. The table has columns for Name, Severity, Policy Status, Entity Category, Scope, and Alert Condition. The first two policies are '15-minutes Average Load High' and '1-minute Average Load High', both with a severity of 'Critical' and a status of 'Disabled'. The remaining policies have various severities and statuses. Below the table, the details for the '15-minutes Average Load High' policy are shown, including its name, severity, status, and a description: 'Trigger alert if 15-minutes Average Load on Management Node is greater than 10'. The policy is currently disabled.

| Name | Severity | Policy Status | Entity Category | Scope | Alert Condition |
|---|----------|---------------|-----------------|-----------------|--|
| 15-minutes Average Load High | Critical | Disabled | Node | Management Node | 15-minutes Average Load is greater than 10 |
| 1-minute Average Load High | Critical | Disabled | Node | Management Node | 1-minute Average Load is greater than 10 |
| Available Root Disk Space Low | Warning | Enabled | Node | Management Node | Available Root (/) Disk Space (MB) is less than 100 |
| Chassis Fan Modules Critical | Critical | Enabled | Chassis | Any Device | Fan Modules Status is Critical |
| Chassis Health Critical | Critical | Enabled | Chassis | Any Device | Health Status is Critical |
| Chassis Health Warning | Warning | Enabled | Chassis | Any Device | Health Status is Warning |
| Chassis Power Supplies Critical | Critical | Enabled | Chassis | Any Device | Power Supplies Status is Critical |
| Clock Skew High | Warning | Enabled | Node | Any Node | Local Clock Offset (second) is greater than 10 |
| CPU Utilization High on Management Node | Critical | Enabled | Node | Management Node | CPU Usage (%) is greater than 70 |
| Current Input Packet Error Rate High | Warning | Enabled | Switchport | Any Device | Current Input Packet Error Rate (%) is greater than 10 |
| Current Input Packet Loss Rate High | Warning | Enabled | Switchport | Any Device | Current Input Packet Loss Rate (%) is greater than 10 |

Policy Name: 15-minutes Average Load High

Summary | **Alert History**

Properties

Name: 15-minutes Average Load High
 Severity: Critical
 Status: Disabled

Description: Trigger alert if 15-minutes Average Load on Management Node is greater than 10
 Entity Category: Node
 Scope: Management Node

Rules

Condition: 15-minutes Average Load is greater than 10

Resource Reporting – Gauge utilization of the cluster

- Historical reports can be generated for
 - Cluster availability
 - Cluster performance and usage
 - Free application licenses

IBM Platform HPC 4.2

Resource Reports

| Name | Type | Summary |
|--------------------------------|--------|--|
| Free Licenses for Applications | Sample | Number of free licenses for a specific feature |
| Node Availability | Sample | Node availability in the system |
| Node Resource Usage | Sample | Resource usage trends for selected nodes |
| System Entitlement | Sample | Total number of nodes in the system |

Report: Free Licenses for Applications

Type: Sample

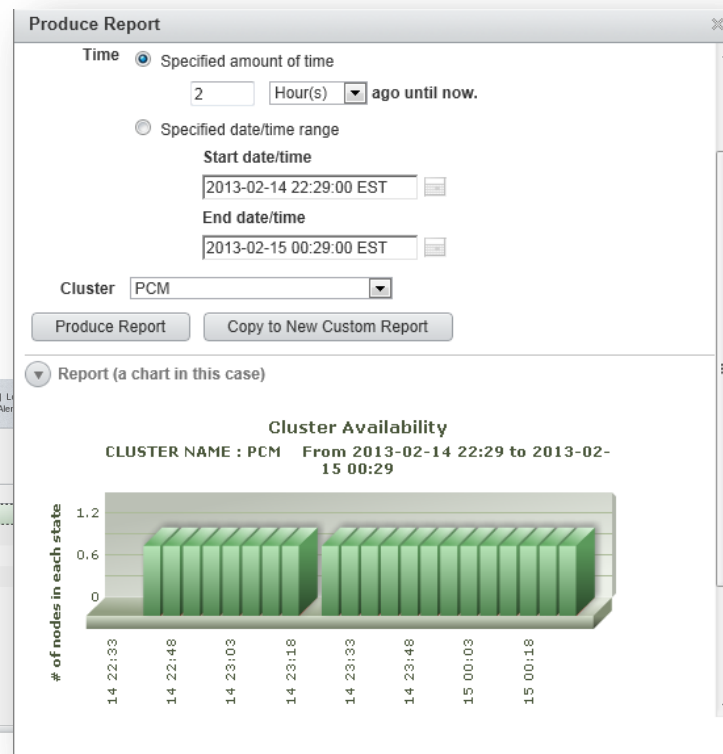
Summary: Number of free licenses for a specific feature

Description: Purpose
This report shows the free license number for selected features over time.

Prerequisite
The license must already be added and configured in order for this report to generate.

Data Description
This report displays the free license number for the selected feature over the specified period of time.
To generate this report, the maximum number of features is 10.

Analysis



Comprehensive HPC Software Stack

| | Products | Client Benefits |
|--------------------------|---|--|
| Systems Management | Platform Cluster Manager – Advanced Edition | <ul style="list-style-type: none"> .Ease of Use: web portal .Customizable: admin productivity .Faster time to system productivity .Robust monitoring |
| Application Runtime | PE Runtime Ed ESSL / PESSL | <ul style="list-style-type: none"> .Optimized Parallel Runtime .Optimized LAPACK and ScaLAPACK libraries .User controlled workflow support |
| Development Productivity | PE Developer Ed XL Compiler | <ul style="list-style-type: none"> .Modern application development environment using Eclipse .Performance analysis tools to help analyze applications .Optimized compiler for Power |
| Workload Management | Platform LSF | <ul style="list-style-type: none"> .Optimized utilization of resources .Policy, energy and resource aware scheduling .Robust add-on features |
| Data Management | Spectrum Scale HPSS TSM | <ul style="list-style-type: none"> .Scalable/reliable storage for parallel filesystem (GSS) .ILM for transparent migration of data from storage to tape and back |
| Application Environment | Platform Application Center Platform Process Manager | <ul style="list-style-type: none"> .Simplify job submission for repeatable workload: customization .Customizable .Faster time to system productivity |

Platform Computing



Power Systems™ S824L

Power Systems



Power Systems S822L

OpenPOWER™

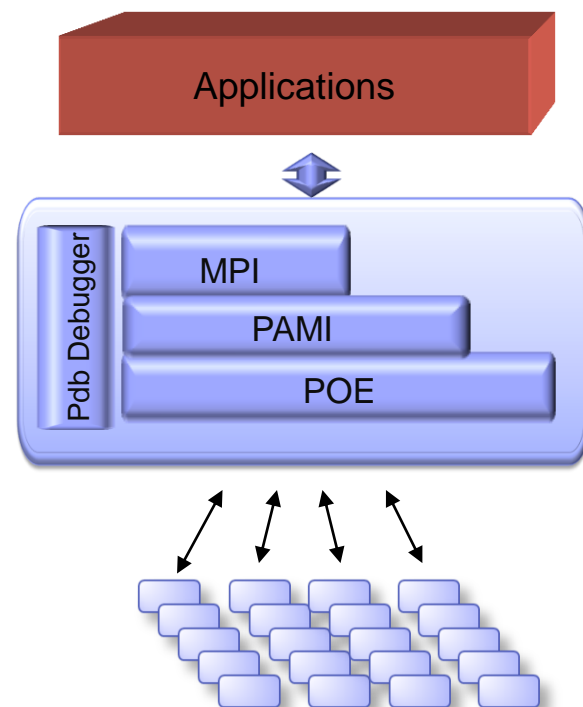
IBM Parallel Environment

High Performance Execution Environment to Take Full Advantage of Scalable Compute Resources

- Parallel Operating Environment (POE) for submitting and managing jobs.
- IBM's MPI and PAMI libraries for communication between parallel tasks.
- A parallel debugger (pdb) for debugging parallel programs.
- IBM High Performance Computing Toolkit for analyzing performance of parallel and serial applications.
- Integrated with LSF to assist in resource management, job submission and node allocation

What's New:

- Ubuntu 14.04.1 Little Endian NV (Non Virtualized)
- MPICH as BASE and collective performance improvements
- MPI 3.0 (via MPICH)
- MPI I/O Improved Performance



Linux on
POWER8
LE support

IBM Platform LSF

Most Complete

- Advanced, feature-rich workload scheduling
- Robust set of add-on features
- Integrated application support

Most Powerful

- Policy & resource-aware scheduling
- Resource consolidation for max performance
- Advanced self-management

Most Scalable

- Thousands of concurrent users & jobs
- Virtualized pool of shared resources
- Flexible control, multiple policies

Best TCO

- Optimal utilization, less infrastructure cost
- Better productivity, faster time to result
- Robust capabilities, administrative productivity

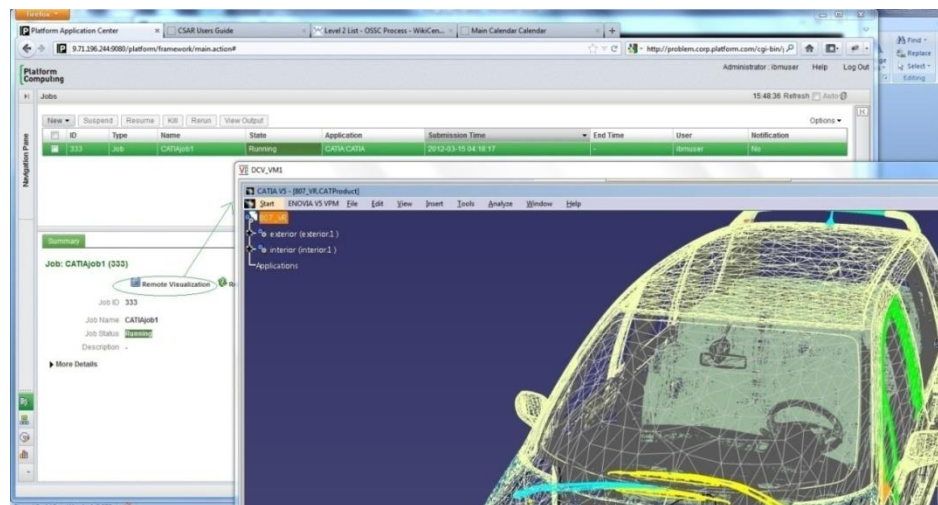


OpenPOWER™

Linux on
POWER8
LE support

IBM Platform Application Center

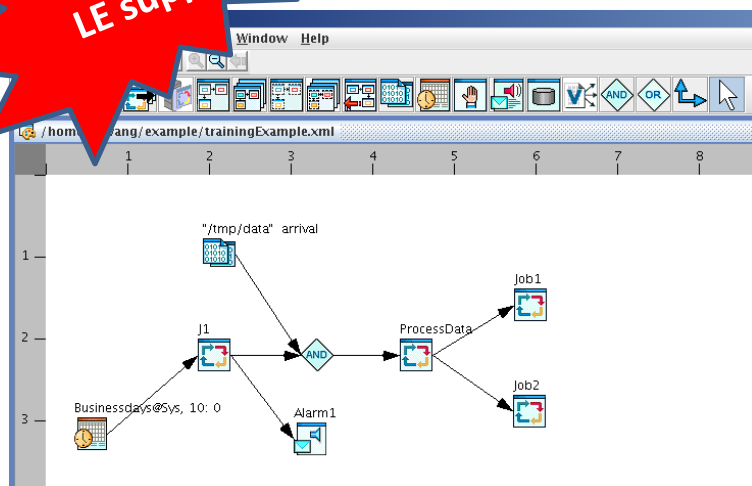
- Increase user productivity with browser based access for job submission, management
- Capture best practices with guided submission (templates)
- Enable access via mobile devices with web services
- Support for 2D/3D remote visualization



OpenPOWER™

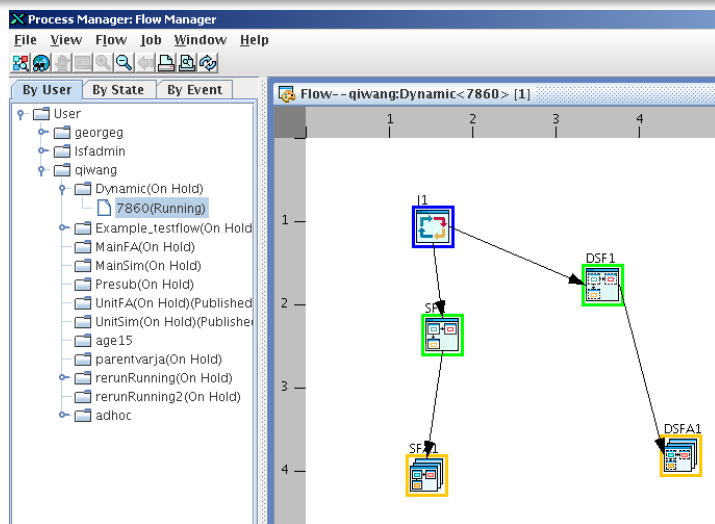
Linux on
POWER8
LE support

IBM Platform Process Manager



Platform Process Manager Flow Editor

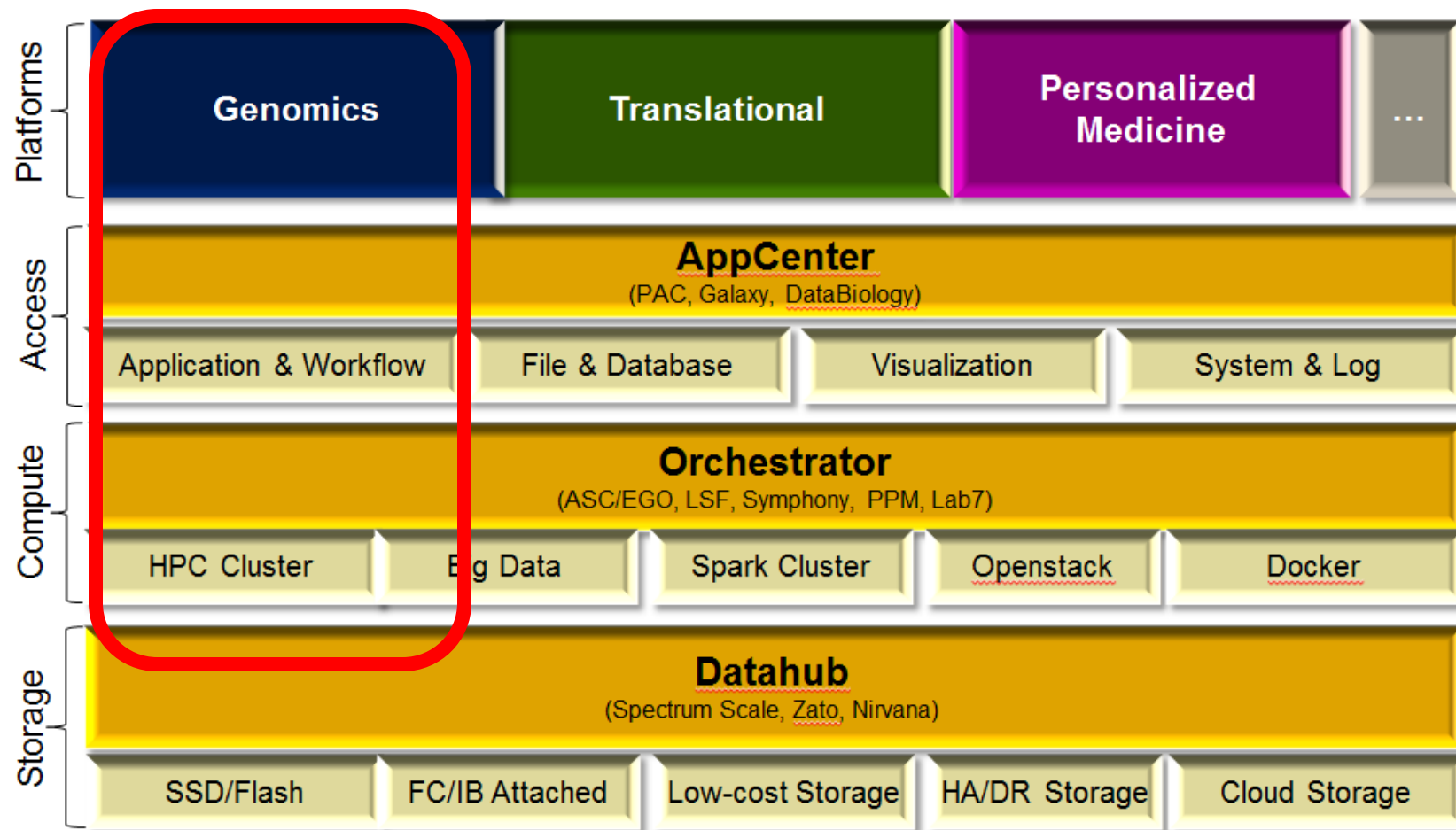
- Intuitive drag-and-drop interface
- Creates self-documenting flows
- Support for sub-flows, job arrays
- Rich error-handling / retry capability
- Save workflows in XML format
- Publish flows directly to Flow Manager



Platform Process Manager Flow Manager

- Manages multiple flows for multiple users and groups simultaneously
- Monitor workflow execution graphically
- Trigger flows automatically through calendar events, the flow manager or the command line.

Genomic medicine – reference architecture



<http://www.powergene.net>

Links

- IBM Platform Computing product information (<https://ibm.biz/BdXBDR>)
- Service Management Connect – Technical Computing Community (<https://ibm.biz/BdFr8R>)
- IBM Knowledge Center (<https://ibm.biz/BdXBDX>)



Contacts

- Development Manager: Jing Li (jingili@cn.ibm.com)
- Product Management: Mehdi Bozzo-Rey (mbozzore@ca.ibm.com)
- Product Marketing: Gabor Samu (gsamu@ca.ibm.com)



Q&A