IBM Platform LSF & PCM-AE
Dynamische Anpassung von HPC Clustern für Sondernutzung und Forschungskollaborationen

ZKI Meeting 2012 - Düsseldorf

Heiko Lehmann
Account Manager
Heiko.lehmann@de.ibm.com

Bernhard Schott
EMEA SE Manager
schott1@de.ibm.com
Progress in operations of dynamic HPC centers

HPC center operations and management is simple and straightforward! Only issue: the users always want something different …

- IBM and Platform Computing
  - Die Erfolgsgeschichte geht weiter
- Herausforderungen im HPC Betrieb
  - *Ein* Cluster für alle – geht das (noch)?
- IBM Platform Cluster Manager – AE
  - Unterstützung von speziellen Anforderungen
  - Isolation: HPC-Cluster on demand
- IBM Platform LSF – Multicluster
  - Standortübergreifende Forschungskollaborationen
  - Erhöhte Anforderungen an Datenschutz /Privacy /EU Data Protection Directive
Platform Computing, Inc.

Platform Clusters, Grids, Clouds Computing

The leader in managing large scale shared environments

- 19 years of profitable growth
- Customers: 9 of the Global 10 largest companies
- 2,500 of the world’s most demanding client organizations
- 6,000,000 CPUs under management
- Headquarters in Toronto, Canada
- 500+ professionals working across 13 global centers
- World Class Global Support
- Strong Partnerships with Dell, Intel, Microsoft, Red Hat VMWare
### IBM Platform Computing Partner Ecosystem

#### Strategic Partners
- CRAY
- FUJITSU
- IBM
- Microsoft
- Dell
- HP
- Intel
- SAS

#### Premier Partners
- GemStone
- Synopsys
- NICE
- The MathWorks
- Mentor Graphics
- Cadence
- ANSYS
- Bull
- APMW
- The BioTeam
- Vinetech
- Wipro
- Acer
- Schlumberger
- MSC Software
- Calypso
- SYBASE
- SGI
- SCi
e
- NEC
- Mellanox Technologies
- LSTC
- Wolfram Research
- Vigen
- Dassault Systemes

#### Select Partners
Herausforderungen im HPC Betrieb

HPC center operations and management is simple and straightforward! Only issue: the users always want something different …

- Contradicting requirements

  1) Offer only one OS, patch level, software stack: RHEL6.2
  2) Support important applications

- Example: User with big funding / budget / important project:
  “My app needs SUSE10/SLC5.7/CentOS4.6 – why don’t you support me?”

  Clash!
Herausforderungen im HPC Betrieb

- Contradicting requirements

3) Keep operations and support costs low, users cannot modify systems
4) Validation of new versus legacy application(s) (versions)

- Example: Life-Science user needs to validate old SW version against new version: controlled changes in app, libraries, patches and switch them back and forth. User needs root rights to do the changes himself.

Big issue!
Herausforderungen im HPC Betrieb

HPC center operations and management is simple and straight forward! Only issue: the users always want something different …

- Contradicting requirements

5) Need for compute power $\rightarrow$ Oversubscription factor 5 (JSC)
6) Keep certain usages contained & isolated

- Example: Divide compute power (and budget) into multiple bigger and smaller islands due to non-technical constraints – efficiency is dropping since resources cannot longer be shared.

Not satisfactory!
Herausforderungen im HPC Betrieb

HPC center operations and management is simple and straightforward! Only issue: the users always want something different …

- Contradicting requirements

7) Apply security patches in timely fashion to minimize risk of exploits.
8) Keep application environments unchanged

- Example: After security patches were applied, support calls are exploding; My application does not run anymore! You cannot change “my” system without consulting me!

Drop security for peace with your users?
Herausforderungen im HPC Betrieb

- Contradicting requirements

9) Collaborate with other research institutions.
10) Keep users productive – complex Grid solutions are inadequate

- Example: Funding was granted for research collaboration with Site B and Site C. How can we build a collaborative “single system image” that is easy to setup and maintain and easy to use? How to prevent disruptive changes in our HPC environment?

Create collaboration without disruption for my own site?
Progress in operations of dynamic HPC centers

- Resources, users and applications requirements are heterogeneous and ever changing
- Special users and collaborations are kicking the boundaries.
- HPC center operations need to become - and can be straightforward & flexible

Solution: flexible system management by IBM Platform Cluster Manager Advanced Edition

IBM PCM-AE
Building a dynamic HPC infrastructure with IBM Platform Cluster Manager Advanced Edition

Jeff Karmiol
Sr Product Manager
jkarmiol@ca.ibm.com
IBM Platform Cluster Manager Advanced Edition

- Management software for building HPC/Technical computing clouds
  - Automates the self-service creation, flexing and management of heterogeneous HPC clusters for use by multiple tenants.

- HPC Cloud Foundation
  - Unifies resources within and across data centers
  - Self-service portal and catalog
  - Automated provisioning
  - Dynamic runtime management
  - Single administration pane
  - Reporting and monitoring

- Delivers increased utilization and decreased capital expenditures and utilization
Capabilities for HPC Clouds

- **Deploy HPC Clusters On Demand**
  - Rapidly deploy multiple heterogeneous HPC clusters on a shared hardware pool
  - Self-service allows users to request a custom cluster, specifying size, type and timeframe
  - A multi-tenant environment provides separation between user groups and customized user experiences.

- **Cluster “Flexing”**
  - Dynamically grow and shrink (flex up/down) the size of a deployed cluster based on workload demand, calendar and sharing policies.
  - Share hardware across clusters by rapidly re-provisioning to meet the infrastructure needs
Main Advantages

- **Multi-platform HPC cluster support**
  - IBM Platform LSF, IBM Platform Symphony, SGE, PBS, …
  - All your clustering technologies are available on-demand through your service catalog

- **On-demand HPC self-service cluster provisioning**
  - Get the resources when you want them.
  - Self-service and provisioning eliminates the need to submit a request and wait for someone to act.

- **HPC Cluster scaling**
  - Support expansion and shrinking of clusters as needed over time.

- **Rapid HPC cluster provisioning**
  - Get the clusters you need, in minutes, instead of hours and days.

- **Get HPC physical and virtual resources**
  - Choose the resources type to match the job.

- **Multi-tenancy HPC support**
  - User separation, different service catalogs, resource limits for sharing, per account reporting

- **Shared HPC Cloud Services Enablement**
  - Extend to external “as-a-service” (e.g. IBM SmartCloud) for peak demand
Platform Cluster Manager Advanced Edition

User/Admin Self-Service Clusters
- Web User Interface
- CLI

Reporting & Accounting
- Allocation Engine: Reservations, Resource Sharing Plans, Priorities, Interconnect Awareness
- SLA Manager: Policy-Driven Growing & Shrinking of Clusters

Resource Integrations
- Physical Host Provisioning
- Hypervisor Management: Linux, KVM
- Customer Specific Provisioning Adaptor
- Virtualization Adaptor: CCS/SCP 2.1 (Target 4Q12)

Operational Management Clusters & Machines
- Cluster Support: Define, Deploy, Flex HPC Clusters

ISV Integrations
- Process Manager
- License Scheduler
- Application Center
- IBM Platform LSF
- Map-Reduce
- IBM Platform Symphony

Other Integrations
- SGE
- PBS Pro
- Hadoop
- IBM MPI
- PE Developer Edition
User Roles

Cloud Admin
- Build resource pools
- Define service catalog & deliver services
- User and groups, quotas

Account Managers
- Manage users and subgroups
- Prioritize & allocate resource quotas
- Manage catalog

End Users
- Request & use resources subject to quota
- Monitor deployed resources
Manage All Machine Resources

Manage all available Physical & Virtual Machines

Direct machine control
Manage Multiple Clusters

Manage all deployed clusters

View individual machines
Manage Cluster Resources

- Manage all deployed clusters
- Drill down from Cluster to individual machine
- Access Cluster Console (PAC)
Self-Serve Clusters

1. Select
2. Configure
3. Life span
4. Size
5. Submit
6. Deployed
Dynamic Clusters

- Manual
- Load
- Calendar
Define Clusters for Self Service

Repository of predefined cluster building elements

Machine Provision

Pre-Provision Script

Post-Provision Scripts

Cluster Designer
Edit Pre- and Post-Provisioning Scripts

Script in any provisioned language (e.g., Linux shells, Perl, Python ...)

```bash
# provision_compute -
# Install and start LSF Compute host.
#
provision_compute()
{
    LOG "Installing LSF ...
    LSF_MASTER_LIST="$1"
    export LSF_MASTER_LIST

    #unzip the installation package.
    if [ ! -d $DESTINATION_DIR/LSF_INSTALL_PACKAGENAME ]; then
        tar -xzvf $DESTINATION_DIR/$(LSF_INSTALL_PACKAGENAME)_linux_x86_64.tar.Z C $DESTINATION_DIR 2>>$LOG_FILE
    fi

    if [ "$0" = "0" ]; then
        echo "LSF master installation failed. Refer $LOG_FILE on $LOCALMACHINE for detail information." >&2
        return 1
    fi

    #Modify slave config
    sed -i -e '# $LSF_TOP=/usr/local/sbin/LSF_CLUSTER_TOP/" $DESTINATION_DIR/LSF_INSTALL_PACKAGENAME/lsf.config
    sed -i -e '# $LSF_ADMIN_USER="lsfadmin user1 user2" $LSF_ADMIN_USER="" $DESTINATION_DIR/LSF_INSTALL_PACKAGENAME/lsf.config
    sed -i -e '# $LSF_SERVER_HOSTS="host1 host2 host3" $LSF_SERVER_HOSTS="" $DESTINATION_DIR/LSF_INSTALL_PACKAGENAME/lsf.config
    sed -i -e '# $LSF_MASTER_LIST="" $LSF_MASTER_LIST="" $DESTINATION_DIR/LSF_INSTALL_PACKAGENAME/lsf.config

    #Modify lsinstall
    sed -i -e '#show_secret' $DESTINATION_DIR/LSF_INSTALL_PACKAGENAME/lsinstall

    #Modify lsconfig
    sed -i -e '#show_secret' $DESTINATION_DIR/LSF_INSTALL_PACKAGENAME/lsconfig

    echo "LSF master installation complete."

    LOG "Installing LSF ...
    LSF_MASTER_LIST="$1"
    export LSF_MASTER_LIST

    #unzip the installation package.
    if [ ! -d $DESTINATION_DIR/LSF_INSTALL_PACKAGENAME ]; then
        tar -xzvf $DESTINATION_DIR/$(LSF_INSTALL_PACKAGENAME)_linux_x86_64.tar.Z C $DESTINATION_DIR 2>>$LOG_FILE
    fi

    if [ "$0" = "0" ]; then
        echo "LSF master installation failed. Refer $LOG_FILE on $LOCALMACHINE for detail information." >&2
        return 1
    fi

    #Modify slave config
    sed -i -e '# $LSF_TOP=/usr/local/sbin/LSF_CLUSTER_TOP/" $DESTINATION_DIR/LSF_INSTALL_PACKAGENAME/lsf.config
    sed -i -e '# $LSF_ADMIN_USER="lsfadmin user1 user2" $LSF_ADMIN_USER="" $DESTINATION_DIR/LSF_INSTALL_PACKAGENAME/lsf.config
    sed -i -e '# $LSF_SERVER_HOSTS="host1 host2 host3" $LSF_SERVER_HOSTS="" $DESTINATION_DIR/LSF_INSTALL_PACKAGENAME/lsf.config
    sed -i -e '# $LSF_MASTER_LIST="" $LSF_MASTER_LIST="" $DESTINATION_DIR/LSF_INSTALL_PACKAGENAME/lsf.config

    #Modify lsinstall
    sed -i -e '#show_secret' $DESTINATION_DIR/LSF_INSTALL_PACKAGENAME/lsinstall

    #Modify lsconfig
    sed -i -e '#show_secret' $DESTINATION_DIR/LSF_INSTALL_PACKAGENAME/lsconfig

    echo "LSF master installation complete."
}```
Multi-Tenant

- Set Resource Limits
- Multiple Tenants & Nested Accounts
- Nested Accounts
Physical Node Provisioning
### IBM Platform Cluster Manager Advanced Edition

#### Reports

**Report Type:** Allocation Report  
**Report For:** Accounts  
**Account:** All Accounts  
**Date Range:** Custom Dates  
**From:** Apr 01, 2012 EDT  
**To:** May 24, 2012 EDT

#### Allocation Report

Custom Dates (01 Apr 2012 to 25 May 2012)  
For All Accounts

<table>
<thead>
<tr>
<th>Account Name</th>
<th>Cluster Name</th>
<th>Physical Resource (Machine Hours)</th>
<th>Virtual Resource (Memory Hours in GB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AdminAccount</td>
<td>KVM host</td>
<td>Reserved 10.09, On Demand 0.00, Allocated 10.09</td>
<td>Reserved 0.00, On Demand 0.00, Allocated 0.00</td>
</tr>
<tr>
<td>AdminAccount</td>
<td>KVM host</td>
<td>Reserved 555.78, On Demand 0.00, Allocated 555.78</td>
<td>Reserved 0.00, On Demand 0.00, Allocated 0.00</td>
</tr>
<tr>
<td>AdminAccount</td>
<td>LSF Cluster Virtual_Admin_1</td>
<td>Reserved 0.00, On Demand 0.00, Allocated 829.00</td>
<td>Reserved 0.00, On Demand 829.00</td>
</tr>
<tr>
<td>AdminAccount</td>
<td>LSF Cluster Virtual_Admin_2</td>
<td>Reserved 0.00, On Demand 0.00, Allocated 28.45</td>
<td>Reserved 0.00, On Demand 28.45</td>
</tr>
<tr>
<td>AdminAccount</td>
<td>LSF Cluster Virtual_Admin_3</td>
<td>Reserved 0.00, On Demand 0.00, Allocated 23.79</td>
<td>Reserved 0.00, On Demand 23.79</td>
</tr>
<tr>
<td>AdminAccount</td>
<td>RHEL 6.2</td>
<td>Reserved 23.49, On Demand 23.49, Allocated 23.49</td>
<td>Reserved 0.00, On Demand 0.00, Allocated 0.00</td>
</tr>
<tr>
<td>Total for AdminAccount</td>
<td></td>
<td>589.36, 0.00, 881.23</td>
<td>0.00, 881.23</td>
</tr>
<tr>
<td>Total for All Accounts</td>
<td></td>
<td>589.36, 0.00, 881.23</td>
<td>0.00, 881.23</td>
</tr>
</tbody>
</table>
IBM SmartCloud 2.1 Provisioning, cont…
HPC Cloud Client Use Cases
CAE ISV - Dynamic & SaaS HPC Clusters

Background:

- Client is a developer and distributor of Computer Aided Engineering (CAE) software. Their primary focus is computation fluid dynamics.
- Primary I/T environment run on IBM servers with GPFS in an IBM ITS data center in USA.
- Business plan to offer SaaS option
- Workload manager has been SGE

Customer Goals:

- Enhance “foundational” I/T capacity (e.g. servers, storage) for internal and external use with high-capacity
- Establish portal capability for self-service and multi-tenancy with remote visualization
- Extend to external “variable” I/T capacity for an integrated ability to support peak loads
Solution design

- Two additional racks of x86 servers at IBM data center
- Integration with GPFS storage and FDR Infiniband
- **Platform Cluster Manager** **Advanced Edition for dynamic multi-tenancy and extension to additional IaaS capacity**
  - PCM AE development for needed functionality, integration with external variable capacity suppliers
- Platform Application Center (PAC) for portal and core remote visualization
- Platform LSF for workload management
- Platform Computing services
Client Secure Multi-Tenant On-Demand Environment

Client A

Client B

Client C

Client D

VPN

Self-serve portal to request simulation environment

Using PCMAE
Receives requests
Provisions and manages environments

Engineer

PBS Portal

PAC

PAC

Leased resources

LSF

PBS

SGE

LSF

PBS

IBM Data Center

Clients
Request environment
Upload design
Visualize output
Download output
Dynamic HPC Bio Clusters

Background:

- Major Bio-Research Computing Site
- HPC infrastructure has slow network and distributed data store
- One cluster with single job queue (SGE) that cannot meet user requirements
  - Test & Dev (crash & burn) environment
  - Internal and external collaboration
  - Performance
  - Next-generation genomics sequencing creates huge challenge for data
  - System administrator challenges

Customer Goals:

- Develop optimized HPC architecture
- Enable efficient management of resources
- Enable intelligent allocation of workload
- Enhance user experience
Solution: PCM-AE & LSF

- Users
  - Jobs
  - Cluster Request

- Admins

Platform Application Center
- Open Source Apps
- Proprietary Apps

Platform LSF
- Platform Cluster Manager – Advanced Ed.

GPFS
- RHEL 6.2 + KVM
- IBM
- H/W
- xSeries

Platform RTM

Storage
Research Collaboration in high performance and reduced operations effort
Component 1: IBM Platform LSF MultiCluster

Use & manage your assets as a single, virtual computer from anywhere in the world

Autonomous clusters form a Grid – local stability & global sharing
LSF Multi-cluster Architecture

- **Cluster1 master**
- **Cluster2 master**
- **Cluster3 master**

**Job forwarding**

- MC meta-scheduler
- Virtual resources
- Resource lease

**computers**
Component 2: Platform Cluster Manager Advanced Edition

Proactively provide resources as requested by job requirements. Fresh nodes (fresh OS/patches/app/lib or VM) join the LSF cluster.

PCM-AE “understands” job requirements and provides the requested resources.
Use case example: production proof HPC Cloud

Combining the Power of Unified Clusters

With the flexibility of Dynamic resource management
HPC Cloud for research & collaboration

A flexible AND reliable solution with proven industry strength

• Combine Grid style workload distribution with Cloud style resource management
• Platform Dynamic Cluster provides Cloud style flexibility for resources
• Platform LSF MultiCluster to manage workload intelligently across sites, regions, the globe => a Grid

• Data and network (latency!) aware scheduling.
• About 200 Platform customers use MultiCluster in significant deployments, about 60 global Grids.
• Proven solution with stability, performance, reliability
HPC Cloud for research & collaboration

E-Science Grid 2.0?
HPC Cloud collaboration: principle

Consortium partner
HPC Cloud collaboration: isolation & multiple domains
Thank you!
HPC Cloud for research & collaboration

Two Cloud Bursting Strategies

• Invoke AWS VMs, make them join your cluster
• Invoke a Cloud based cluster, let it join LSF-MultiCluster

Cluster stability depends on long distance link => questionable

*solution specific
HPC Cloud for research & collaboration

Two Cloud Bursting Strategies

• Invoke VMs, make them join your cluster

• **Invoke a Cloud based cluster, let it join LSF-MultiCluster**

Cluster stability depends Cloud internal LAN => good!