

---

# Crowbar

Deploying Cloud Benchmarks

Nicholas Wakou

Dell Cloud Computing Solutions

5/09/2012

---



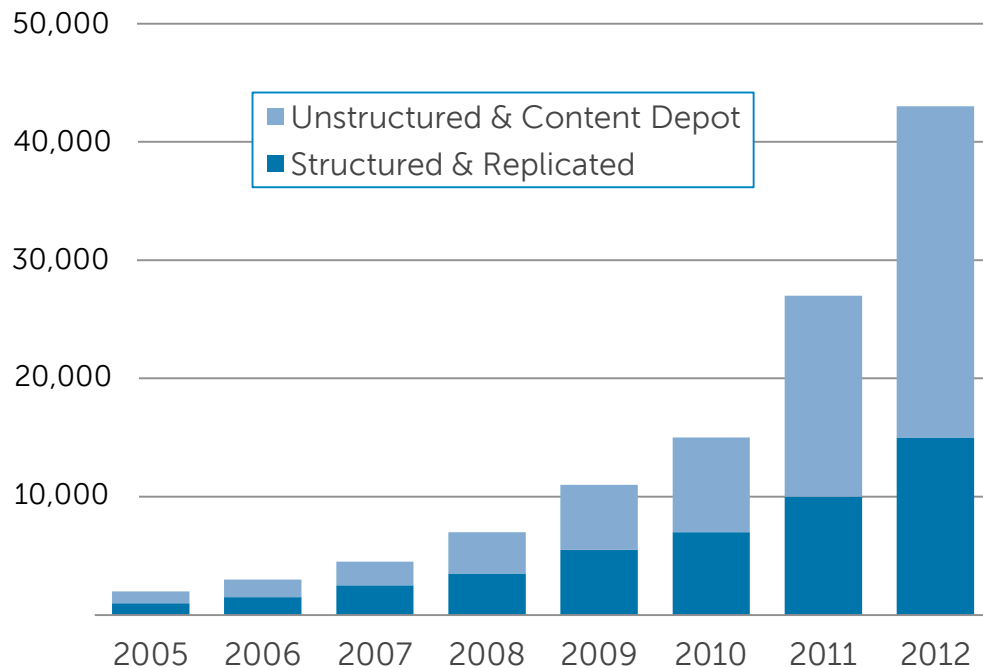
©Dell, 2011

# For the next 5 mins .....

- Big Data realities
- Crowbar
  - ❑ Overview
  - ❑ What is does
  - ❑ High-level components
- Dashboard
- Conclusion

# Reality 1: We are experiencing massive data growth.

## Petabytes



Source: IDC Digital Universe 2009; White Paper, Sponsored by EMC, May 2009

- Every 18 months, non-rich structured and unstructured enterprise data doubles.
- Traditional data warehouse data volumes, while growing, are a small fraction of the data challenge.
- Some 80%<sup>†</sup> of enterprise data resides in other data sources.

<sup>†</sup>Source: Colin White and Claudia Imhoff, "Advanced Analytics and Business Intelligence: Term Abuse?" BeyeNETWORK, May 2010, Web.

# Reality 2: The amount of time required to benchmark big data will grow exponentially

- There is pain in running DW benchmarks
  - Cost
  - Time to Deploy
  - Time to Configure (baselining)
  - Time to Tune
  - Technical expertise
  - Market relevance
- Benchmark SUTs require more resources than production systems
- Big Data benchmark SUTs will require even more resources and need a lot more time to configure.

# Crowbar : Overview

- Open source software framework
- Provides a modular platform for deploying large scale cloud infrastructure
- Automates required installation and configuration tasks from bare metal to deploying big data environments.
- Core capabilities:
  - Hardware configuration – updating and configuring BIOS and BMC boards.
  - Deployment of base operating system.
  - Deployment of cloud/big data components.
  - Providing core network infrastructure services (NTP, DNS, DHCP).
  - Monitoring availability and performance of all deployed components.

# Crowbar – What It Does

- Provides cloud foundation - Bare metal installation, configuration and ongoing configuration management
- Offers catalogue of services/ capabilities (barclamps) for reuse
- Tested on following browsers: Firefox 3.5+, Firefox 4.0, IE 7, Safari 5

Crowbar interfaces:

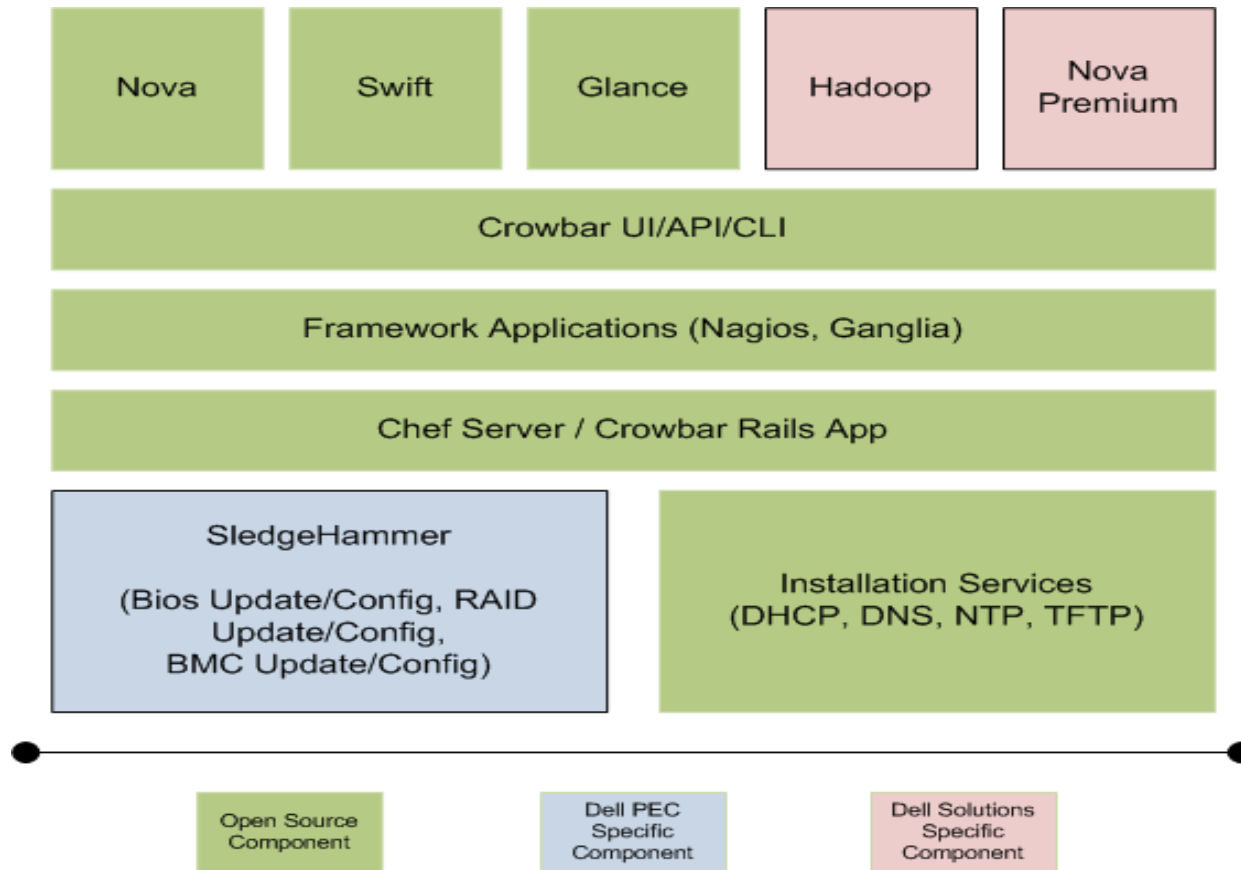
**Nodes**: Core barclamp that does initial configuration of physical or virtual servers to bring them online 'unassigned'. Sets up

- Barclamp framework
- provisioning server,
- base environment and OS,
- network services (ntp, dns, dhcp, VLAN config)
- IP address location service,
- RAID controllers, BIOS and BMC boards.

**Barclamps**: Additional modular software capabilities, i.e. specific roles (titled active proposals) , that can be installed on a selected 'online' nodes. Installs

- Cloudera Manager
- Pig
- Sqoop
- Hive
- Nagios
- Ganglia
- Other components will be added in future

# Crowbar : High-level Components



# Crowbar Dashboard



Dashboard Barclamps Proposals Active Roles Help

There are 12 nodes available in the system.

<b>a4-ba-db-70-f8-74</b>	<b>a4-ba-db-88-92-07</b>
da4-ba-db-17-47-69	d60-eb-69-08-17-86
da4-ba-db-17-44-3f	d60-eb-69-07-de-58
da4-ba-db-14-98-0a	dc8-0a-a9-03-44-46
da4-ba-db-14-70-21	d00-26-9e-cd-e0-c6
da4-ba-db-15-19-b4	dc8-0a-a9-03-44-70
admin	
d00-21-9b-99-9a-4f	

**da4-ba-db-17-44-3f (Edit)** Identify Power On Shutdown Reboot

<b>Full Name</b>	da4-ba-db-17-44-3f.dell.com	<b>Description</b>	Not set
<b>State</b>	Ready	<b>Hardware</b>	PowerEdge R710
<b>Uptime</b>	3 hours 36 minutes 25 seconds	<b>CPU</b>	Intel(R) Xeon(R) CPU E5530 @ 2.40GHz
<b>Switch Name/Port</b>	a4-ba-db-70-f8-74 / 2	<b>Memory</b>	47.26 GB
<b>MAC Address</b>	a4:ba:db:17:44:3f	<b>Disk Drives</b>	8
<b>Allocated</b>	true		

**IP Address**  
bmc: bmc: 192.168.124.173  
nova\_fixed: eth0.500: n/a  
admin: eth0: 192.168.124.91  
[not managed]: eth1:, eth2:, eth3:

**Links**  
IP Mgmt Interface , Nagios , Chef , Ganglia

**Barclamps**  
Bios Default, Deployer Default, Dns Default, Ganglia Default, Ipmi Default, Logging Default, Nagios Default, Network Default, Nova Default, Ntp Default, Provisioner Default, Raid Default

**Roles**  
bios, deployer-client, dns-client, ganglia-client, ipmi-configure, logging-client, nagios-client, network, nova-multi-compute, ntp-client, provisioner-base, raid-configure

Delete Reset Reinstall Hardware Update

<b>a4-ba-db-70-f8-74</b>	<b>a4-ba-db-88-92-07</b>
da4-ba-db-17-47-69	d60-eb-69-08-17-86
da4-ba-db-17-44-3f	d60-eb-69-07-de-58
da4-ba-db-14-98-0a	dc8-0a-a9-03-44-46
da4-ba-db-14-70-21	d00-26-9e-cd-e0-c6
da4-ba-db-15-19-b4	dc8-0a-a9-03-44-70
admin	
d00-21-9b-99-9a-4f	

Provided by CloudEdge Soli

- The main dashboard shows the nodes and if they are ready
  - The columns are the discovered switches and the nodes are under the switch they are connected to
  - Green icon means ready. Red icon means not ready. Next to the switch is a pie chart of the status of the underlying nodes
- Selecting a node will bring up the details for that node. The stats and info about the node is displayed. Selecting a role in the role list will highlighting all nodes that also have that role.





# Conclusion

- Crowbar is open-source - can be customized for benchmark environments
  - Many open-source developers are customizing it for their environments
  - Barclamps can easily be imported into Crowbar.
- Crowbar is being used in both physical and virtual environments
  - Developers test their code in virtual environments before testing on physical
  - Benchmark SUTs can be modeled on virtual environments
  - Performance parameter tuning on virtual clusters
- Remote access and management possible with Crowbar
- Crowbar reduces Time to Deploy (TtD)
  - Saves money
  - Ease of benchmarking
  - Repeatable and Reusable