



HPC ON WALL ST OPENSTACK AND BIG DATA

Brent Holden
Chief Field Architect, Eastern US
April 2014

OPENSTACK FOR HIGH PUT-THROUGH DATA

WORKLOADS ARE EVOLVING



TRADITIONAL WORKLOADS

- Typically resides on a single large Virtual Machine
- Cannot tolerate any downtime
- Needs expensive high availability tools found in VMware vSphere
- Application scales up rather than out

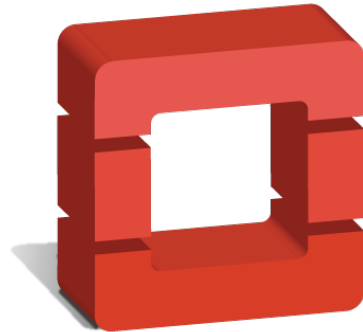
CLOUD WORKLOADS

- Workload resides on multiple Virtual Machines
- Tolerates VM failure – if one fails, another quickly replaces it
- Fault tolerance often built into workload
- Application scales out rather than up

VIRTUAL MACHINE WORKLOAD TYPES

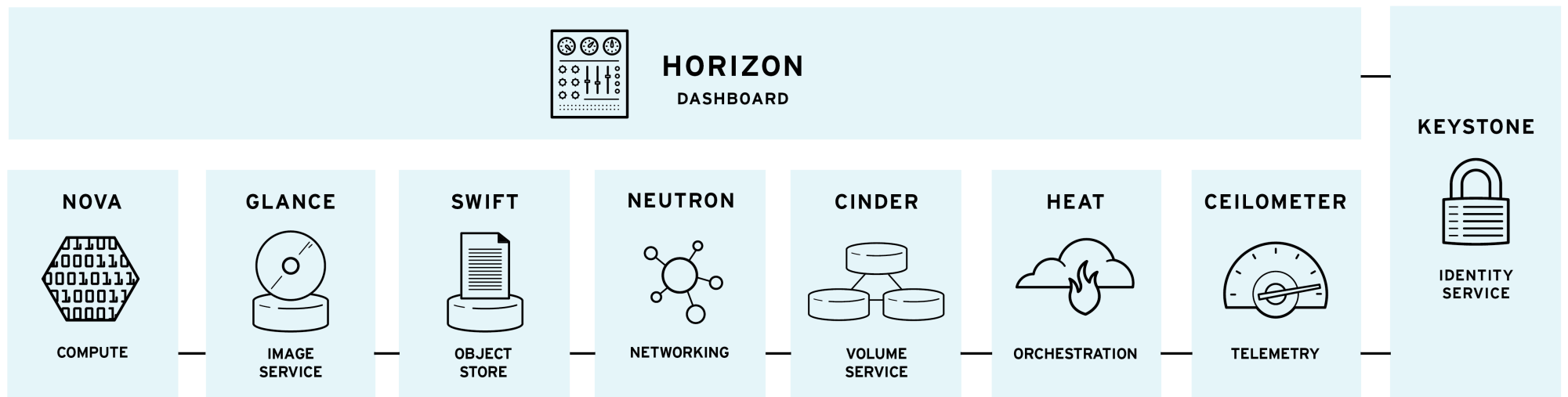
| TRADITIONAL (RHEV/VMWARE) | CLOUD (OpenStack) | MIXED/HYBRID |
|------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Big stateful VM | Small stateless VMs | Combination of Traditional and Cloud VMs to provide application. Database may be hosted on traditional workloads, web front-end and logic layers on cloud workloads. |
| 1 Application → 1 VM | 1 Application → Many VMs | |
| Lifecycle in years | Lifecycle hours to months | |
| Scale up (VM gets bigger) | Scale out (add VMs) | |
| Not designed to tolerate failure of VM, so you need features that keep VMs up | If a VM dies, application kills it and creates a new one, app stays up | |
| Application SLA requires enterprise virtualization features (migration, HA, etc.) to keep applications available | Application SLA requires adding/removing VM instances to application cloud to maintain application availability | |

WHAT IS OPENSTACK?



openstack™
CLOUD SOFTWARE

OPENSTACK ARCHITECTURE



OPST0005

- Modular architecture
- Designed to easily scale out
- Based on a set of core services

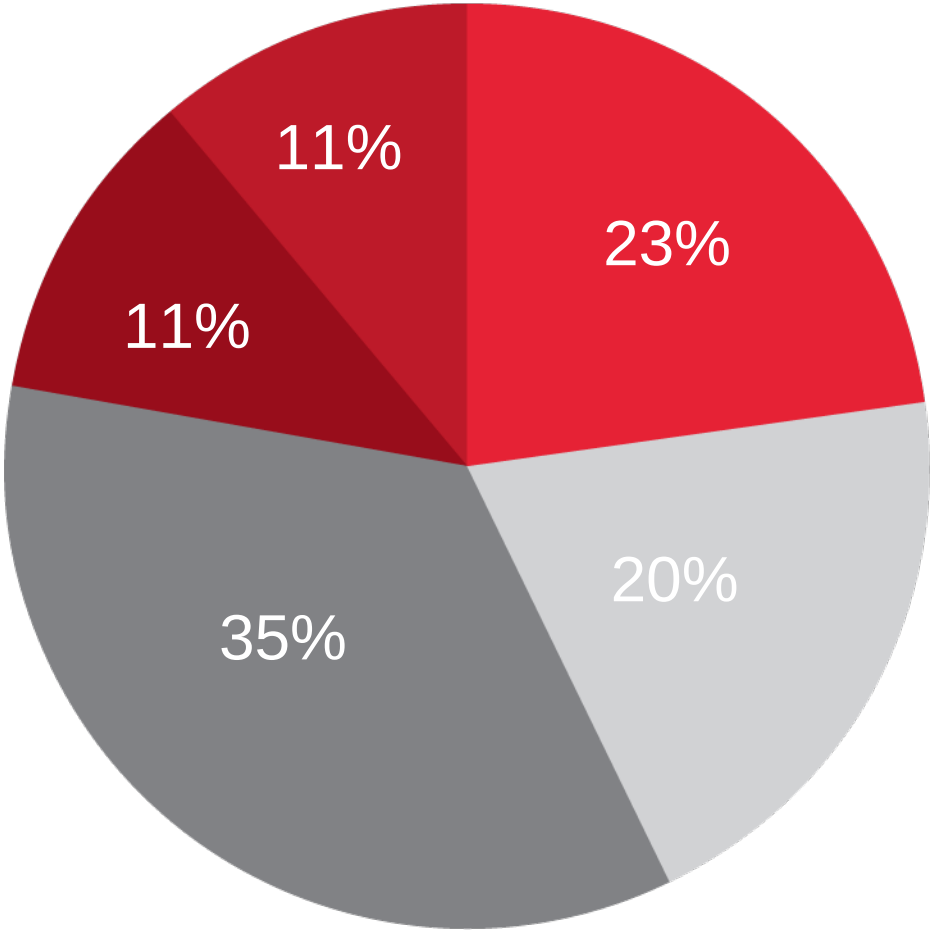
IS OPENSTACK A “CLOUD OPERATING SYSTEM”?

WHY OPENSTACK?

WHY OPENSTACK

- Brings public cloud-like capabilities into your datacenter
- Provides massive on-demand scalability
 - 1,000's → 10,000's of VMs
- It's OPEN!
 - Provides flexibility to customize and interoperate
- Community development = higher “feature velocity”
 - Features and functions you need, faster to market over proprietary software

WHO'S USING OPENSTACK?

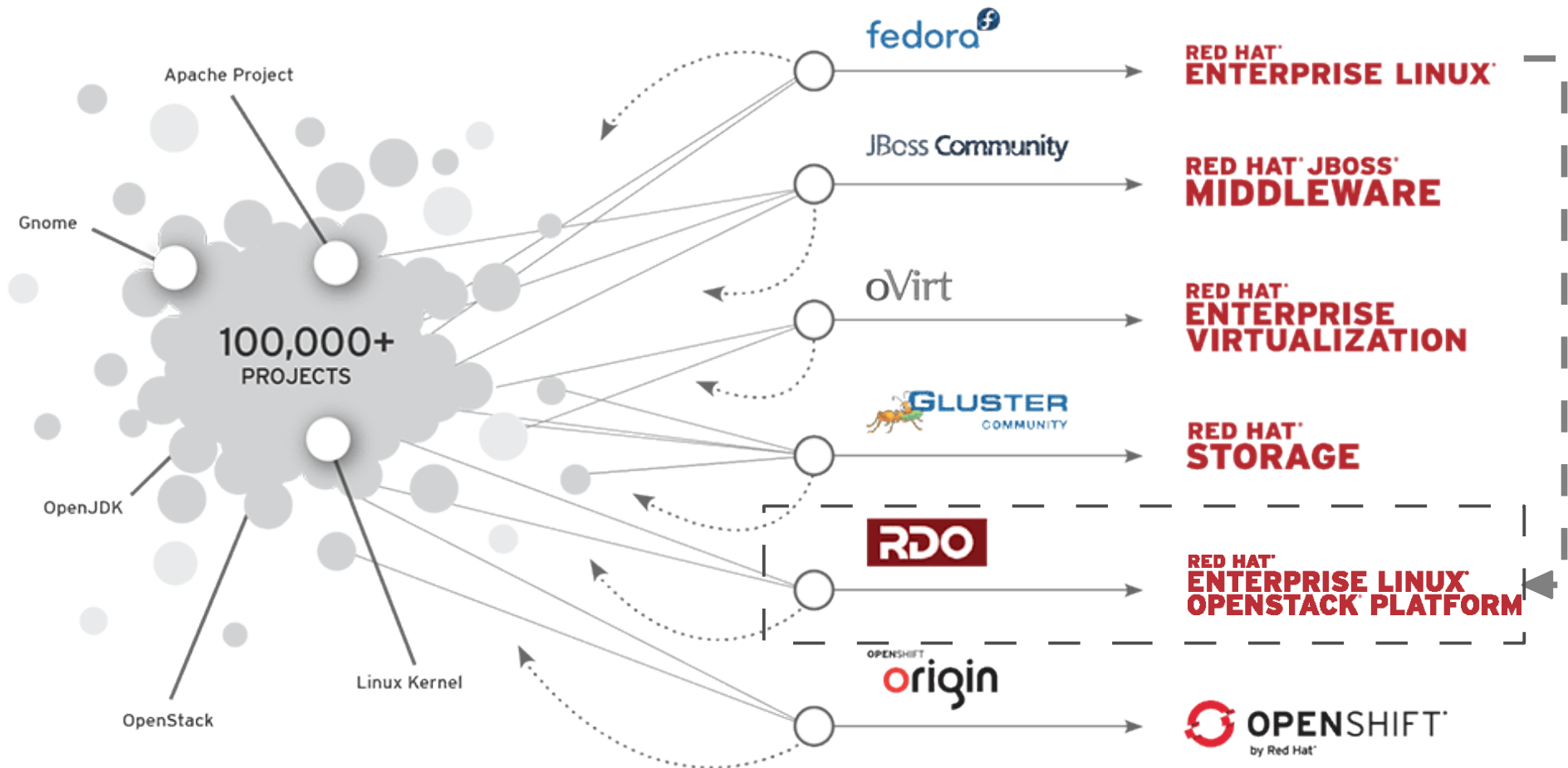


- 65% of respondents have deployed or are planning to deploy OpenStack®
- 62% of respondents plan to use OpenStack for a private cloud infrastructure
- 42% of respondents need OpenStack for massive scalability

- Have deployed
- Currently in deployment
- Will deploy within 24 months
- Plan to deploy, but no timeline
- Other

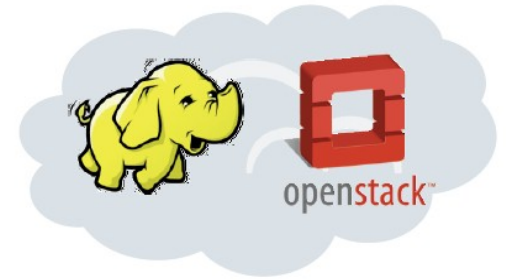
Source: IDG Research – CIO QuickPulse May 2013

RED HAT LEADS THROUGH OPEN INNOVATION



HOW DOES OPENSTACK DO BIG DATA?

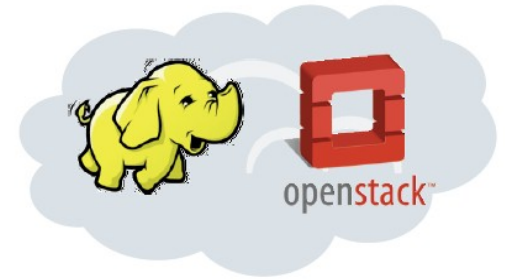
PROJECT SAHARA



MISSION STATEMENT

To provide the OpenStack community with an open, cutting edge, performant and scalable data processing stack and associated management interfaces

PROJECT SAHARA

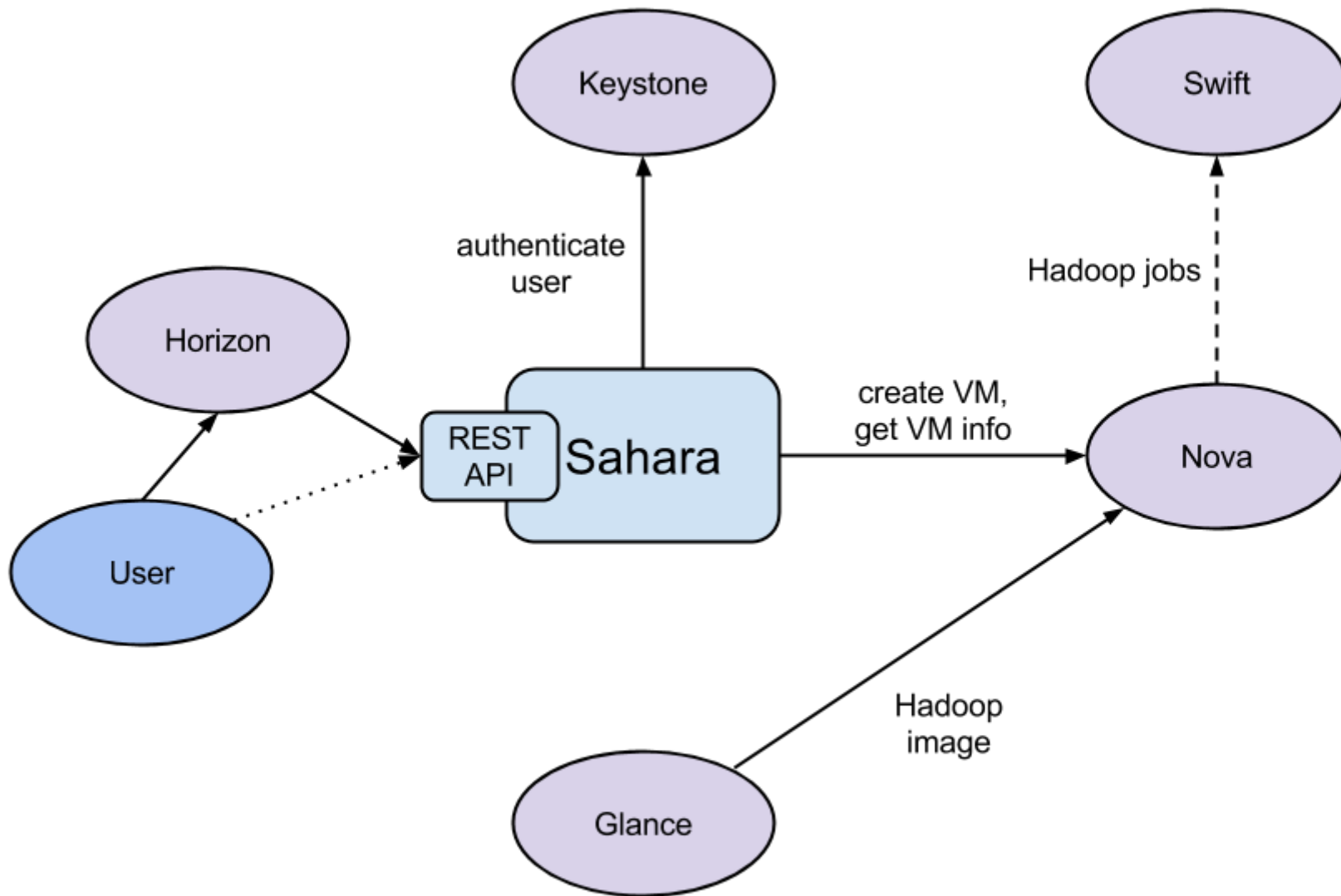


- Formerly known as project 'Savanna'
- Hadoop cluster and workload management
 - Construct the cluster and manage its lifecycle
 - Workflow for big data processing resembling AWS EMR
- Designed to be elastic (scale up/down)
- Designed with a plug-in architecture for multiple vendors
- Provides Analytics-as-a-Service
- *Hadoop is a validation of the OpenStack model*

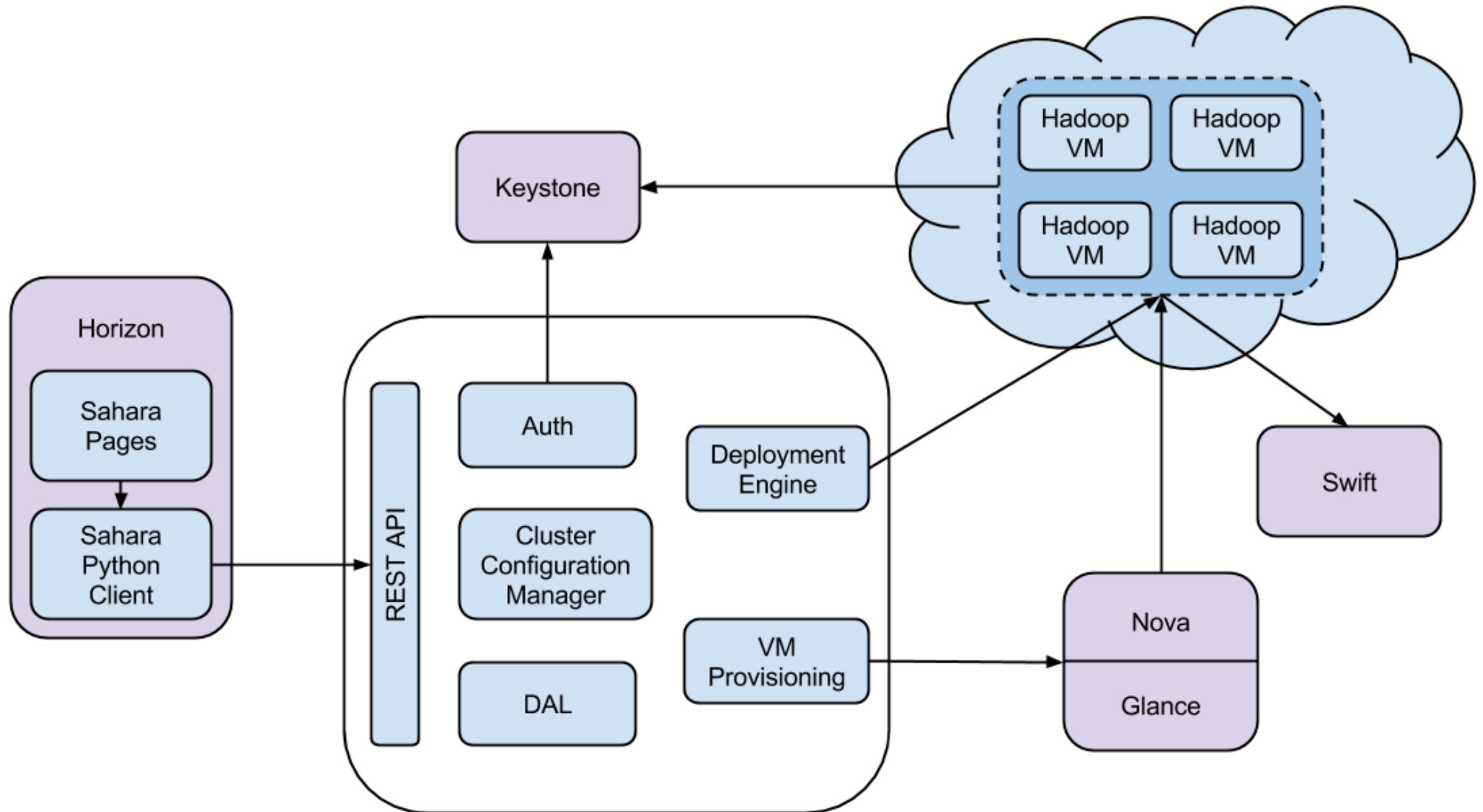
PROJECT SAHARA

- **Why does Hadoop on OpenStack make sense?**
 - **Very popular on AWS for dev/test workloads**
 - **Impractical to move lots of data onto public cloud**
 - **Scale out architecture with fault tolerance**
 - **Works well with “cloud” storage (no dependency on block)**
 - **Provides a place for standard Hadoop operations**
 - **New workloads carry less baggage, legacy requirements**

SAHARA INTEGRATION



SAHARA ARCHITECTURE



SAHARA IN ACTION

The screenshot shows the OpenStack Elastic Hadoop dashboard. The browser address bar indicates the URL `172.18.79.139:9001/project/hadoop/`. The dashboard header includes the OpenStack logo and the text "Elastic Hadoop". A green notification box in the top right corner displays the message: "Success: Create cluster completed successfully." Below the header, there are two main sections: "Hadoop Clusters" and "Node Templates".

Hadoop Clusters

| <input type="checkbox"/> | Cluster Name | Node Templates | Base Image | Status | Nodes Count | Actions |
|--------------------------|--------------|--------------------------------|---------------|----------|-------------|---------------------|
| <input type="checkbox"/> | hadoop | j_n.small: 1 tt_dn.small: 5 | hadoop-img-02 | Starting | 0 | Edit Cluster More ^ |

Deploying 1 item

Node Templates

| <input type="checkbox"/> | Node template name | Node Type | Flavor name | Actions |
|--------------------------|--------------------|-----------|-------------|----------------------|
| <input type="checkbox"/> | j_n.small | JT+NN | m1.small | Edit Template More ^ |
| <input type="checkbox"/> | j_n.medium | JT+NN | m1.medium | Edit Template More ^ |
| <input type="checkbox"/> | j_small | JT | m1.small | Edit Template More ^ |
| <input type="checkbox"/> | j_medium | JT | m1.medium | Edit Template More ^ |
| <input type="checkbox"/> | m_small | NN | m1.small | Edit Template More ^ |

PROJECT SAHARA

- To learn more, go to:
 - <https://wiki.openstack.org/wiki/Savanna>
- Due to graduate with the Icehouse release!

We've been **OPEN** all along.

It's in our DNA.

It's not lip service. Or cloud washing.

Open is what we do. And how we do it.

With every step forward, Red Hat opens another layer of the technology stack.

Cloud is the next step—the next open innovation.

THANK YOU

TRADEMARK STATEMENTS

Copyright © 2014 Red Hat, Inc. Red Hat, Red Hat Enterprise Linux, the Shadowman logo, JBoss, MetaMatrix, and RHCE are trademarks of Red Hat, Inc., registered in the U.S. and other countries. Linux® is the registered trademark of Linus Torvalds in the U.S. and other countries.

The OpenStack™ Word Mark and OpenStack Logo are either registered trademarks / service marks or trademarks / service marks of the OpenStack Foundation, in the United States and other countries and are used with the OpenStack Foundation's permission. We are not affiliated with, endorsed or sponsored by the OpenStack Foundation or the OpenStack community.

