

TUT5605: Deploying an elastic Hadoop cluster

Alejandro Bonilla

Sales Engineer

abonilla@suse.com



Agenda

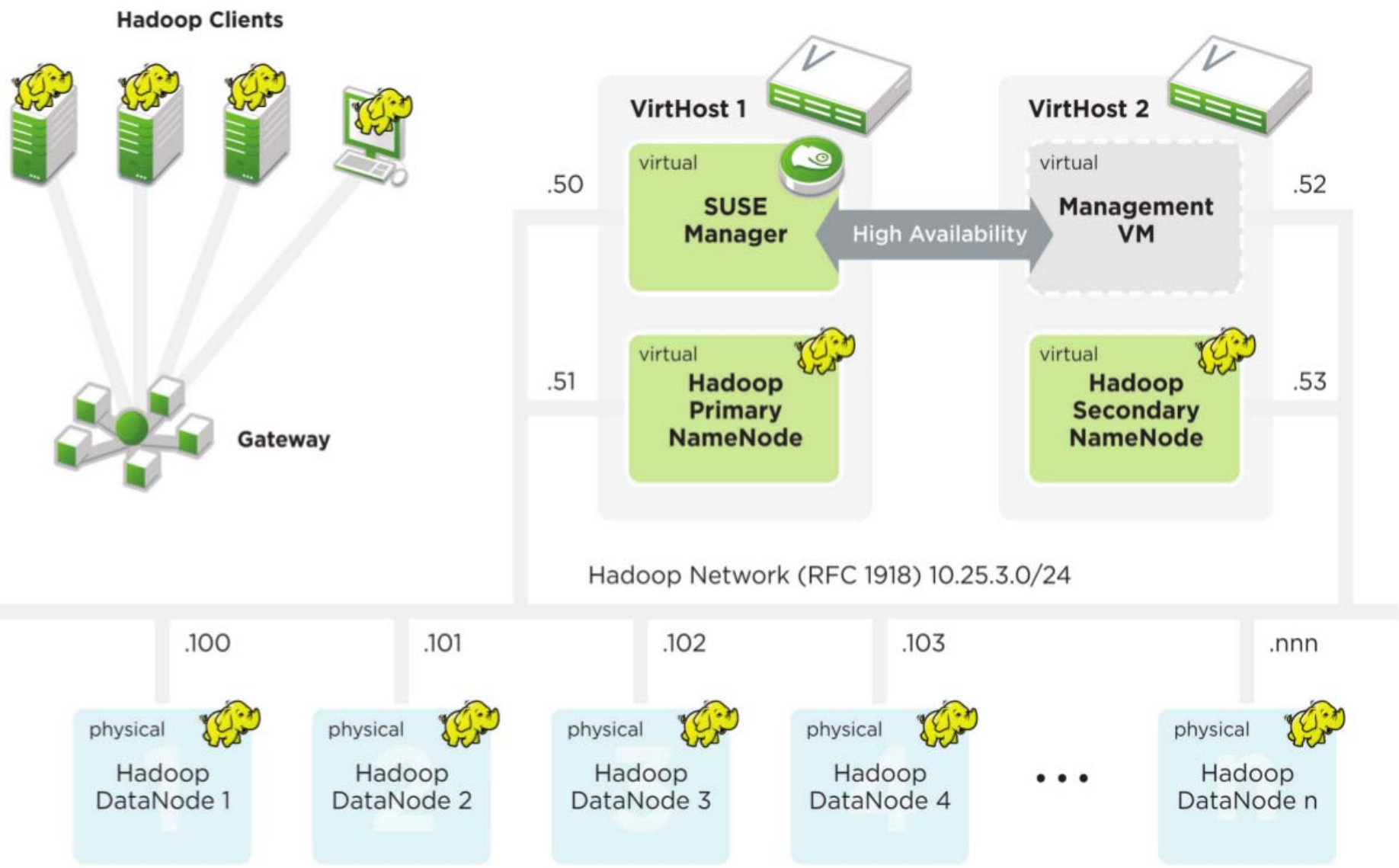
- Overview
- Manual Deployment
- Orchestration
 - Generic workload autoscaling
- Sahara
 - Dedicated for Hadoop Ecosystem
- QA

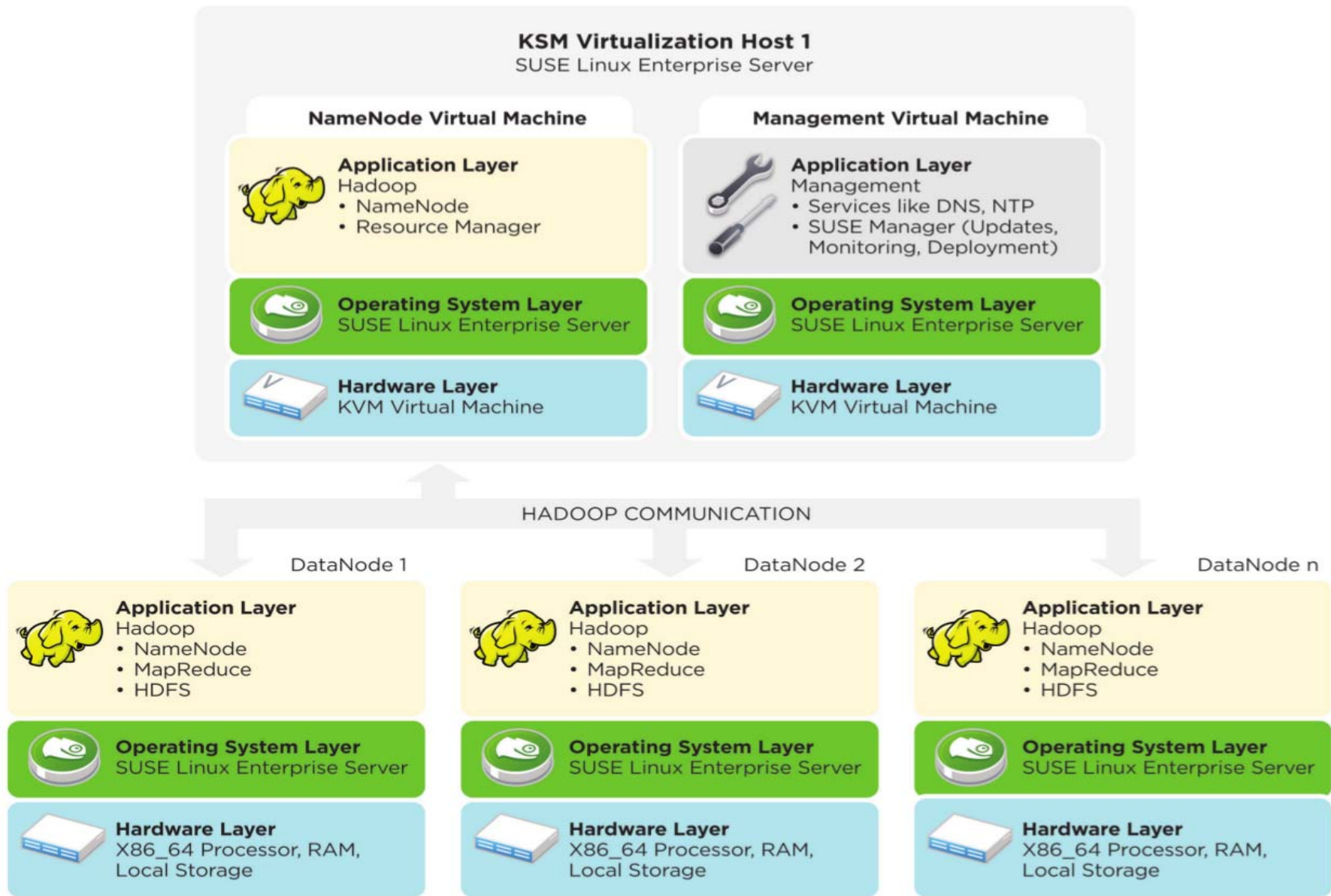
Overview

Hadoop

- Framework that allows for the distributed processing of large data sets across clusters of computers
- The NameNode stores all metadata
- Datanodes carry the Data blocks and execute jobs locally.

- Store data that is relevant in bulk.
- Applications are adjusted to better mine unstructured data.
- Analyze marketing campaigns, fuel efficiency, weather, sequences and predictions.





SUSE Cloud



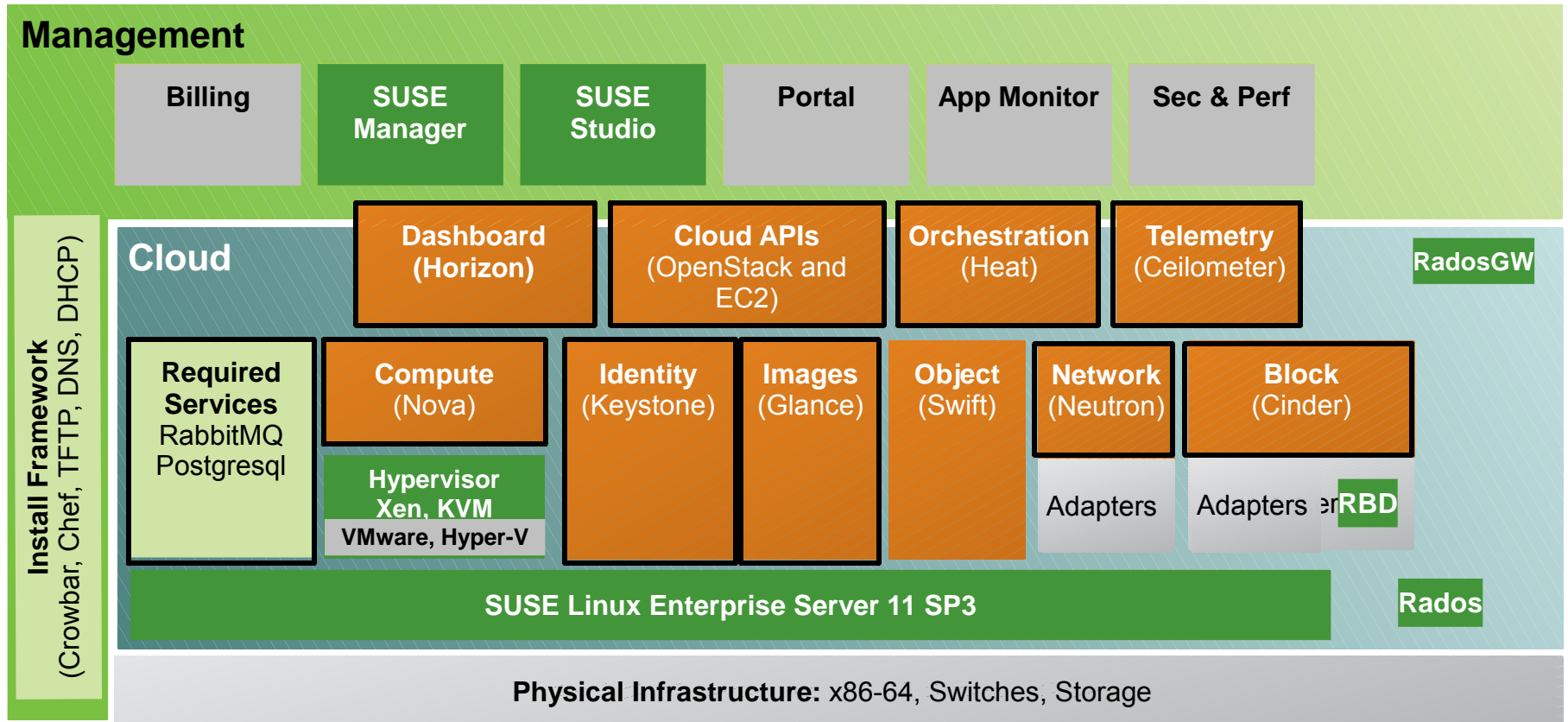
Enterprise OpenStack distribution that rapidly deploys and easily manages highly available, mixed hypervisor IaaS Clouds

- Increase business agility
- Economically scale IT capabilities
- Easily deliver future innovations



Overview

SUSE® Cloud 4



- OpenStack Icehouse
- SUSE Cloud Adds
- SUSE Product
- Ceph
- Highly Available Services
- Partner Solutions

Overview

Why on OpenStack?

- Advanced Resource Management, allocation.
- Flexibility, and ease on Deployment – Virtual or Physical.
- Easy for App Dev testing.
- KVM proves higher performance as it matures.
- After hours analytics of locally stored data / Resource re-utilization.
- Growing Community projects based on OpenStack.

Manual Deployment

Manual Deployment

- Requires a base image with Hadoop already pre-configured.
- Static Setup
- Appropriate for a mature Application Cycle stage.
- May require the Management framework “secret sauce” of Hadoop vendors like Cloudera, Hortonworks and MapR.
- Basic flexibility.

Orchestration - Heat

Orchestration

Heat

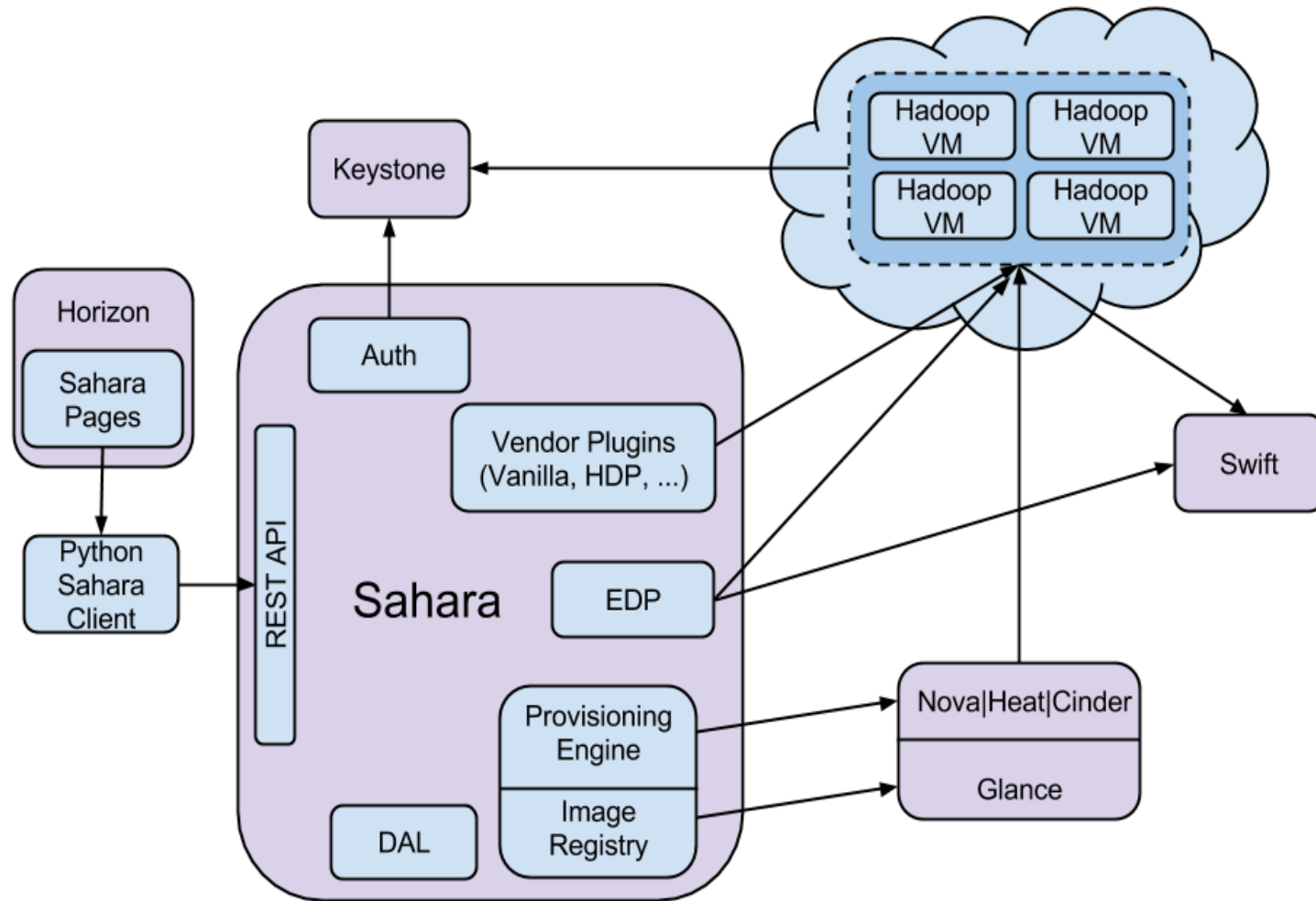
- Implements an orchestration engine to launch multiple composite cloud applications based on templates
- Native Resource Management - servers, floating ips, volumes, security groups, users
- Provides an autoscaling service that integrates with Ceilometer
- Requires a base image with the Hadoop components
- Actions can be triggered based on resource utilization, including auto scale-out

Sahara

Elastic Data Processing facility - EDP

- Sahara's Elastic Data Processing facility allows the execution of jobs on clusters created from Sahara. EDP supports:
 - Hive, Pig, MapReduce, MapReduce.Streaming and Java job types on Hadoop clusters
 - Spark jobs on Spark standalone clusters
 - Access and Storage of job binaries/output in Swift or Sahara's own database
 - Configuration of jobs at submission time
 - Execution of jobs on existing clusters

Sahara Architecture



Demo!

Thank you.







Corporate Headquarters
Maxfeldstrasse 5
90409 Nuremberg
Germany

+49 911 740 53 0 (Worldwide)
www.suse.com

Join us on:
www.opensuse.org

Unpublished Work of SUSE LLC. All Rights Reserved.

This work is an unpublished work and contains confidential, proprietary and trade secret information of SUSE LLC. Access to this work is restricted to SUSE employees who have a need to know to perform tasks within the scope of their assignments. No part of this work may be practiced, performed, copied, distributed, revised, modified, translated, abridged, condensed, expanded, collected, or adapted without the prior written consent of SUSE. Any use or exploitation of this work without authorization could subject the perpetrator to criminal and civil liability.

General Disclaimer

This document is not to be construed as a promise by any participating company to develop, deliver, or market a product. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions. SUSE makes no representations or warranties with respect to the contents of this document, and specifically disclaims any express or implied warranties of merchantability or fitness for any particular purpose. The development, release, and timing of features or functionality described for SUSE products remains at the sole discretion of SUSE. Further, SUSE reserves the right to revise this document and to make changes to its content, at any time, without obligation to notify any person or entity of such revisions or changes. All SUSE marks referenced in this presentation are trademarks or registered trademarks of Novell, Inc. in the United States and other countries. All third-party trademarks are the property of their respective owners.

