SUSE
We adapt. You succeed.

Infraestructura definida por software para el BigData empresarial

Juan Herrera - juan.herrera@suse.com
Noviembre 2017

SUSE

25 YEARS

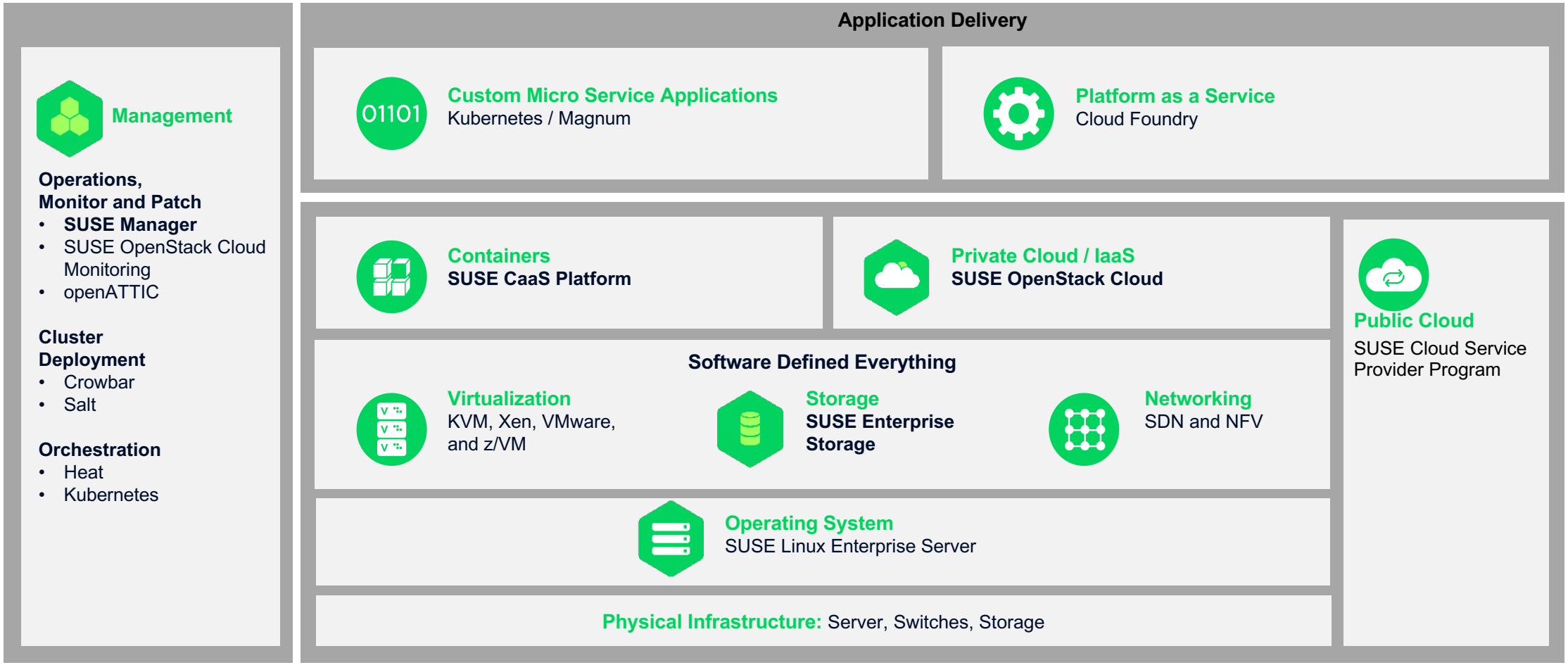
AGENDA

1. SUSE Today
2. BigData Features
3. Management
4. Cloud & Compute
5. Storage
6. ISV
7. IHV
8. SUMMARY

SUSE Today

SUSE Software-defined Infrastructure solutions

An open, flexible infrastructure approach



Product Portfolio

Server and Desktop

- SUSE Linux Enterprise Server
- SUSE Linux Enterprise Server for System z and LinuxONE
- SUSE Linux Enterprise Server for POWER
- SUSE Linux Enterprise Server for ARM
- SUSE Linux Enterprise Server for SAP Applications
- SUSE Linux Enterprise Server for High Performance Computing
- SUSE Linux Enterprise Real Time
- SUSE Linux Enterprise Server with Expanded Support
- SUSE Linux Enterprise Point of Service
- SUSE Linux Enterprise Desktop

Server Extensions

- SUSE Linux Enterprise High Availability Extension
- GEO Clustering for SUSE Linux Enterprise High Availability Extension
- SUSE Linux Enterprise Workstation Extension
- SUSE Linux Enterprise Virtual Machine Driver Pack
- Long Term Service Pack Support
- SUSE Linux Enterprise Live Patching

Cloud, Containers, Storage and Management

- SUSE OpenStack Cloud
- SUSE Enterprise Storage
- SUSE Manager
- SUSE Manager Management Pack for Microsoft System Center
- SUSE Containers as a Service Platform
- SUSE Platform as a Service[1]
- SUSE Studio

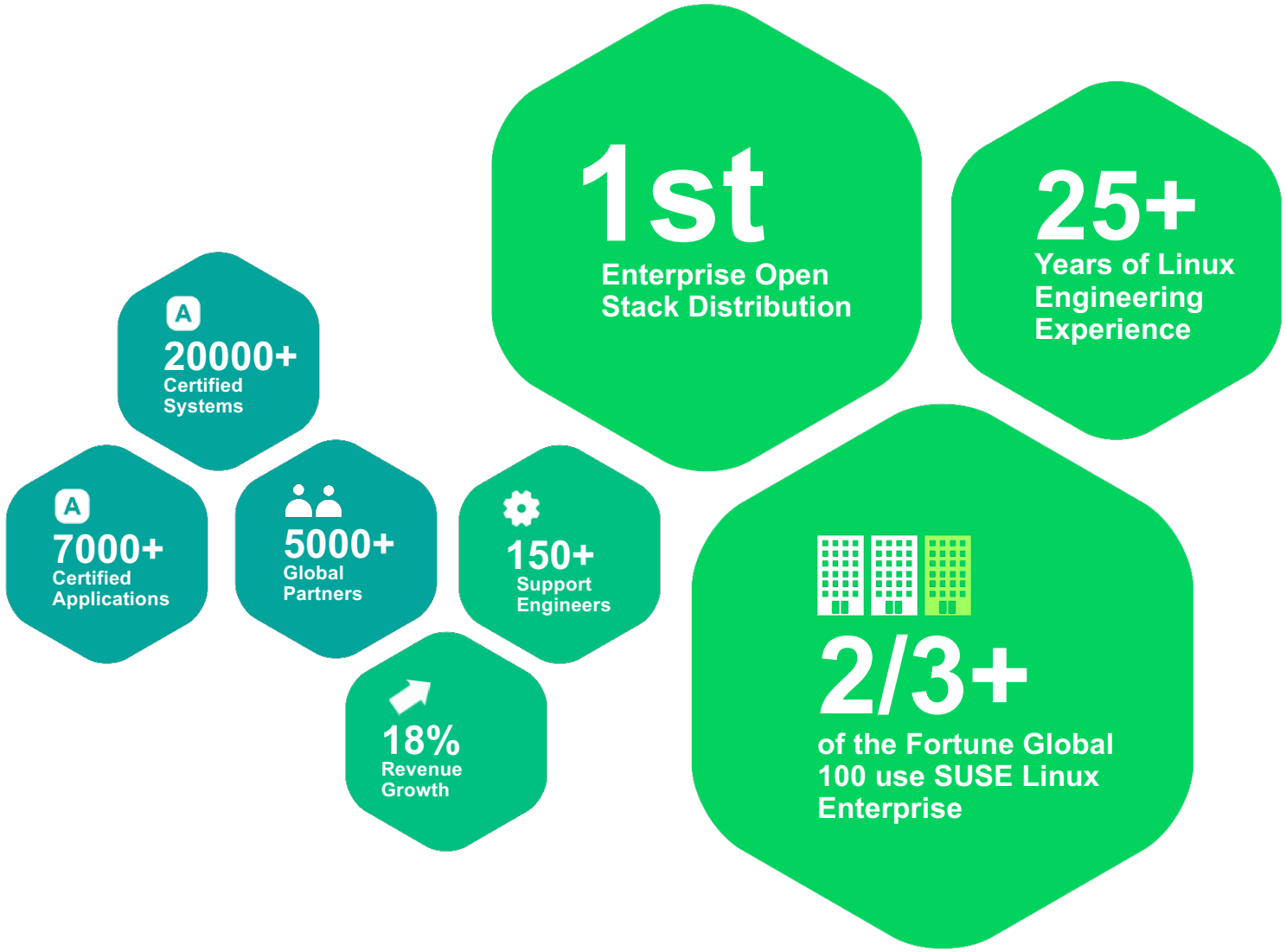
[1] **Q1 2018**

Leading Technology Innovation

SUSE has been the first to:

- Develop enterprise Linux on SAP HANA, AWS and Azure public cloud
- Lead development of the commercial Linux market by delivering the first commercially supported Linux distribution
- Allow instant rollback of operating system changes
- Pioneer continuous availability through live patching for mission-critical systems, including SAP HANA environments
- Deliver a Linux high availability solution that supports geographic mirroring with a broad set of redundancy configurations
- Champion for simplified single system Linux configuration and management (Yast)
- Deliver the first commercially supported OpenStack distribution
- Bring an innovative approach to simplify the deployment of configurable infrastructure (OBS)
- Give consistent support on multiple system architectures by using a common code base
- Provide efficient multiple systems software and asset management built on leading open source technology such as Salt
- Facilitate DevOps adoption through inclusion of Docker technology in SUSE Linux Enterprise Server
- Create the Portus project to simplify and secure management of Docker registries

SUSE at a Glance



Where SUSE Leads

15+ Mainframe Linux

Over 15 years of mainframe Linux market share leadership

4/5 Linux in Finance

4 out of 5 of the world's largest banks use SUSE Linux Enterprise

70% SAP on Linux

70% of all SAP applications running on Linux run on SUSE

95% HANA systems based on SUSE

80% Linux in Large Enterprise

Over 80% of the Fortune Global 50 are active SUSE Customers

9/10 Linux in Aerospace

9 out of 10 of the largest aerospace companies rely on SUSE

x10 Linux in Telecom

10 of the largest telecommunications carriers rely on SUSE

7/10 Linux in Pharma

7 out of 10 of the largest pharmaceutical companies use SUSE Linux Enterprise

7/10 Linux in Retail

7 out of 10 of the largest retailers in the U.S. are active SUSE customers

x10 Linux in Automotive

10 of the largest global automobile mfgs. are active SUSE customers

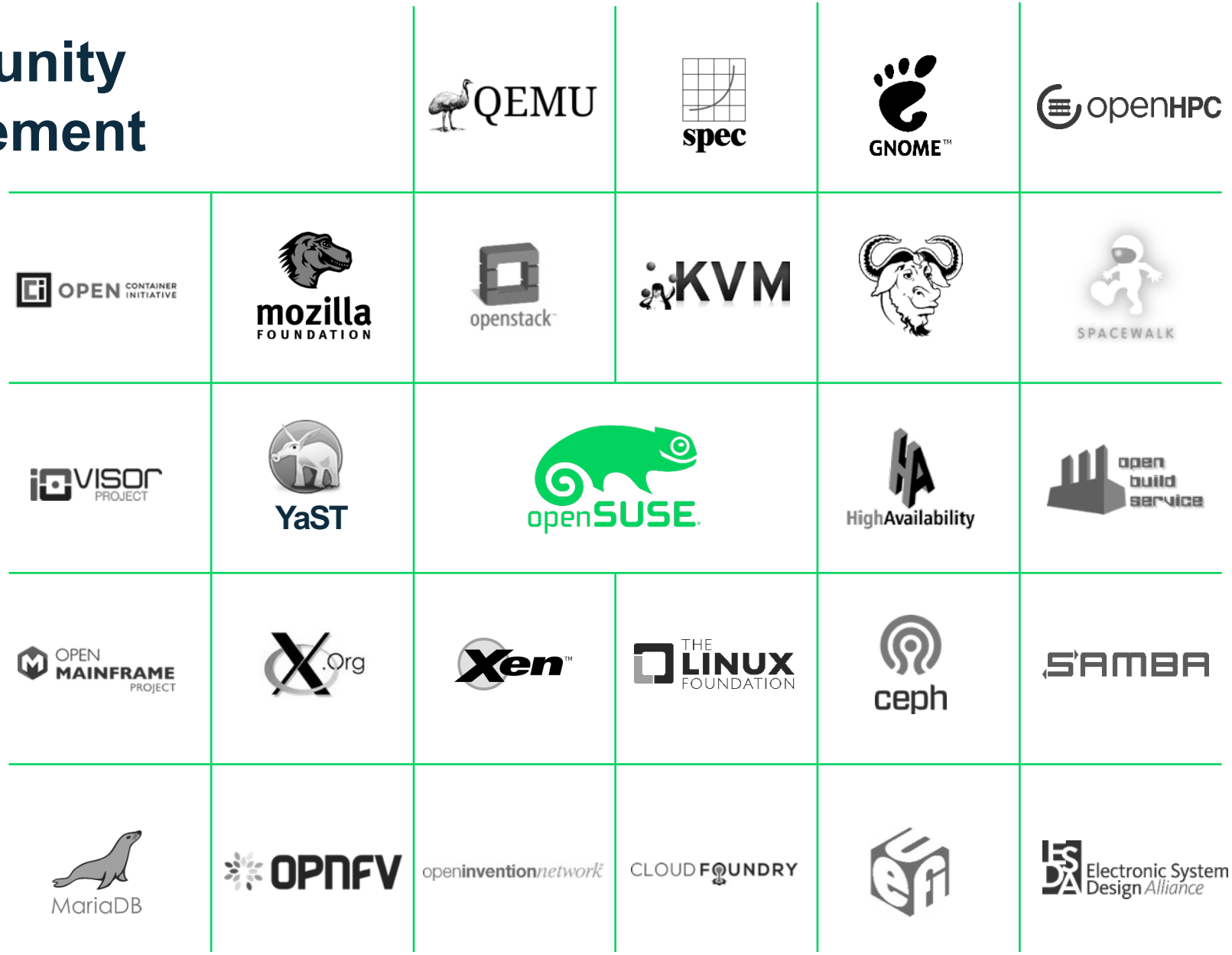
50% Linux in HPC

Half of the world's 20 largest super computers run on SUSE

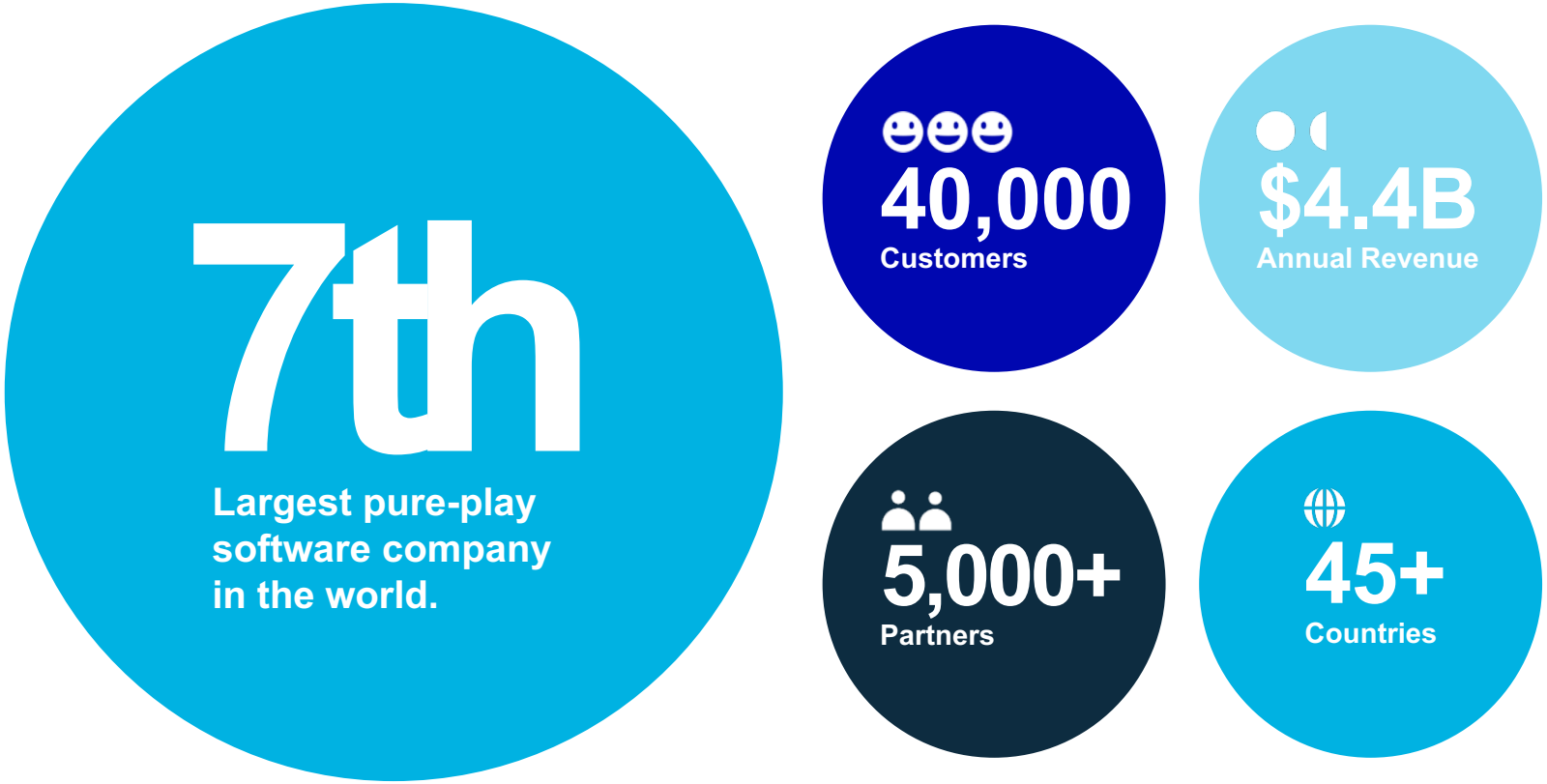
7/10 Linux in Manufacturing

7 out of 10 world's largest manufacturers use SUSE Linux Enterprise

Community Involvement

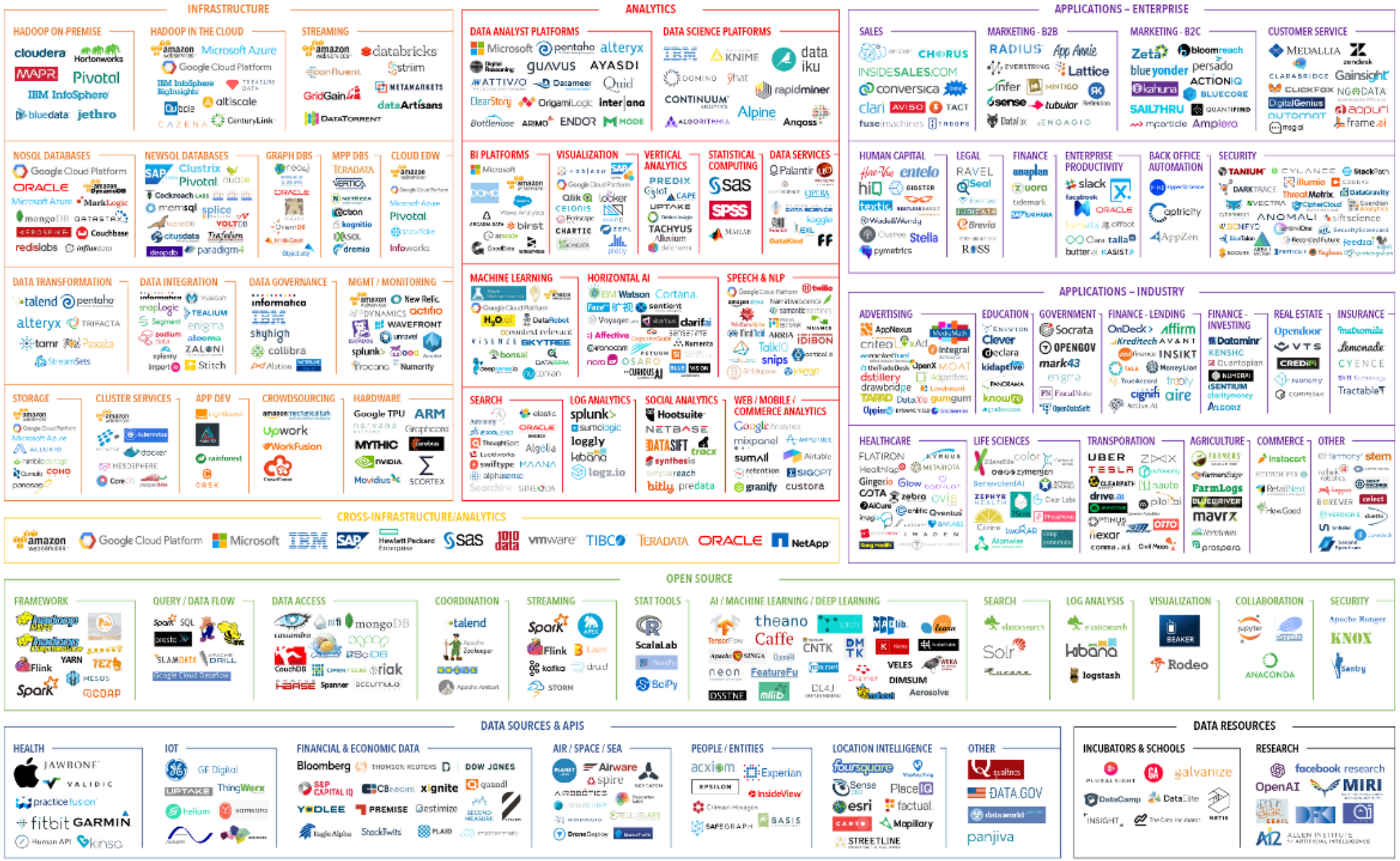


Micro Focus International



Why SUSE for BigData?

BIG DATA LANDSCAPE 2017



Why SUSE for BigData?

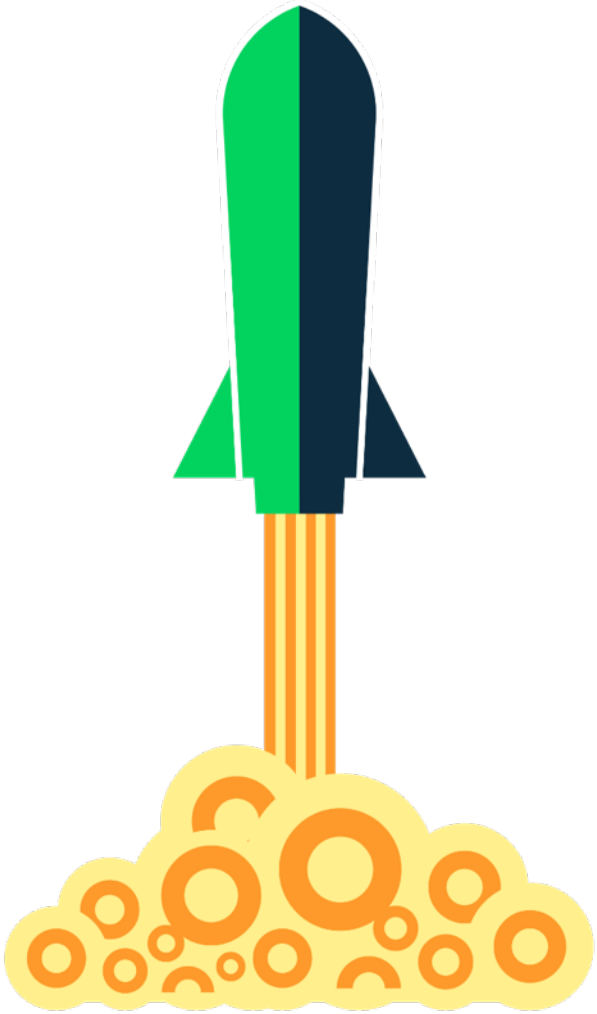
1. SUSE Linux Enterprise Server 12 has features that optimize Big Data deployments
2. Systems management tools for full control of deployment infrastructure
3. SUSE provides perfect-match provisioning architectures for compute, container and storage resources
4. SUSE has partnered with leading Big Data ISVs to ensure SLES is a supported OS
5. SUSE is working with leading IHVs on reference architectures and integrated solutions for Big Data

1. BigData Features

SUSE Linux Enterprise Server is the base

SUSE Linux Enterprise Server as a BigData base OS is:

- Robust and Reliable
- Optimized for massively data-intensive workloads.
- Largest lifecycle in the market
- Easy to install and configure
- Cloud ready



OS Requirements for BigData

BigData ready OS must provide:

High Availability

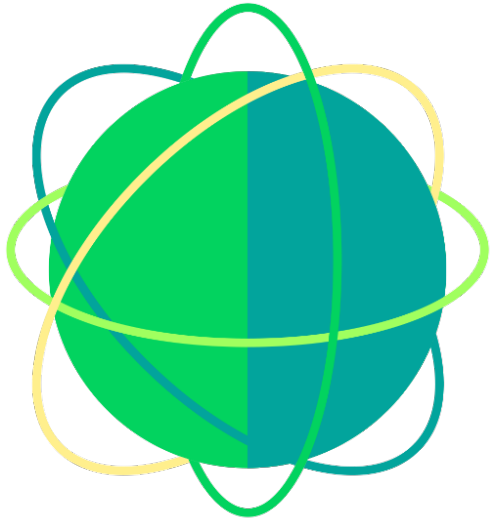
- Control & job scheduling nodes must be highly available

Feature Set

- Scalability, performance tuning, security, data access

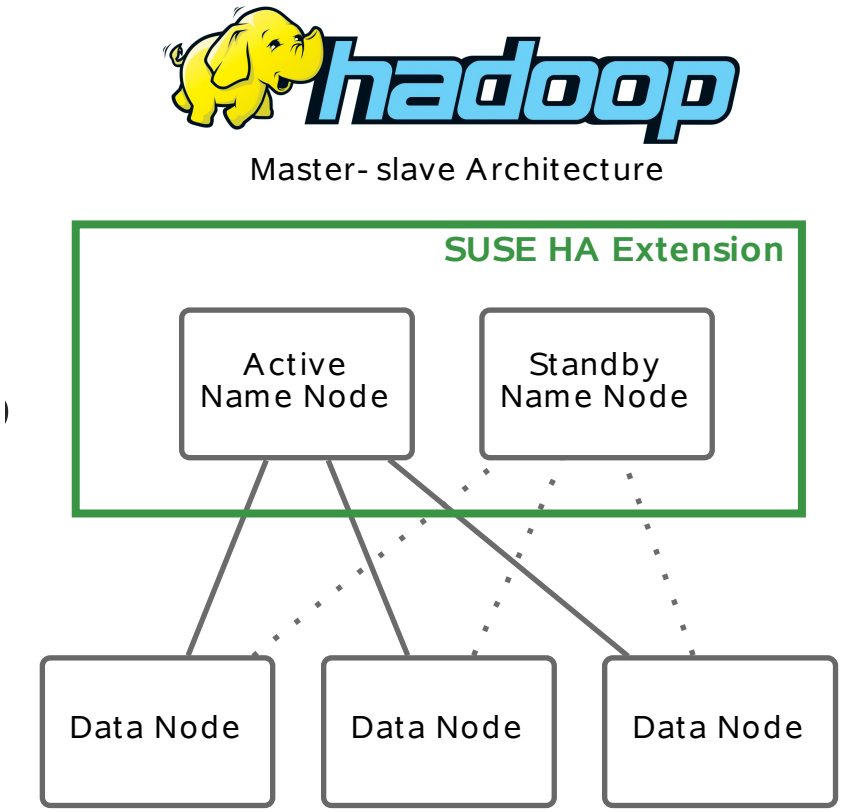
Support

- Maintenance, security updates, critical issues



SUSE Linux High Availability

- Most up-to-date open-source HA stack
- Protects against Hadoop Name Node single point of failure
- Nodes can be physical, virtual, or mixed
- All HA components included (Filesystem, etc.)



SLES 12 Big Data Features

• 4.0 kernel improves performance

- I/O Throttling limits number of I/O streams to avoid throughput delays Improved support for hardware based RAS capabilities
  Scheduler and memory management optimizations
  Faster, more powerful control groups for resource isolation
- Linux Container support for soft partitioning of large physical systems
- Transparent Huge Pages
- kGraft Live Kernel Patching

• Scalability

- CPUs (8192), Filesystem and Storage (XFS 8 EiB), Memory (1 PiB, 64TiB certified)

• Security

- AppArmor proactively protects OS/apps from external/internal threats CC/OSPP EAL 4+ , FIPS 140-2 Certified

• Architecture Support

- x86, x86_64, Itanium, IBM POWER, IBM z/Architecture and ARM64

2. Managing BigData Environments

Big Data Management Considerations

Large-scale deployment -> from 100's or 1000's of nodes

Traditional workload assignment methods don't scale

Distributed systems installation & management

Consistent node installation is important

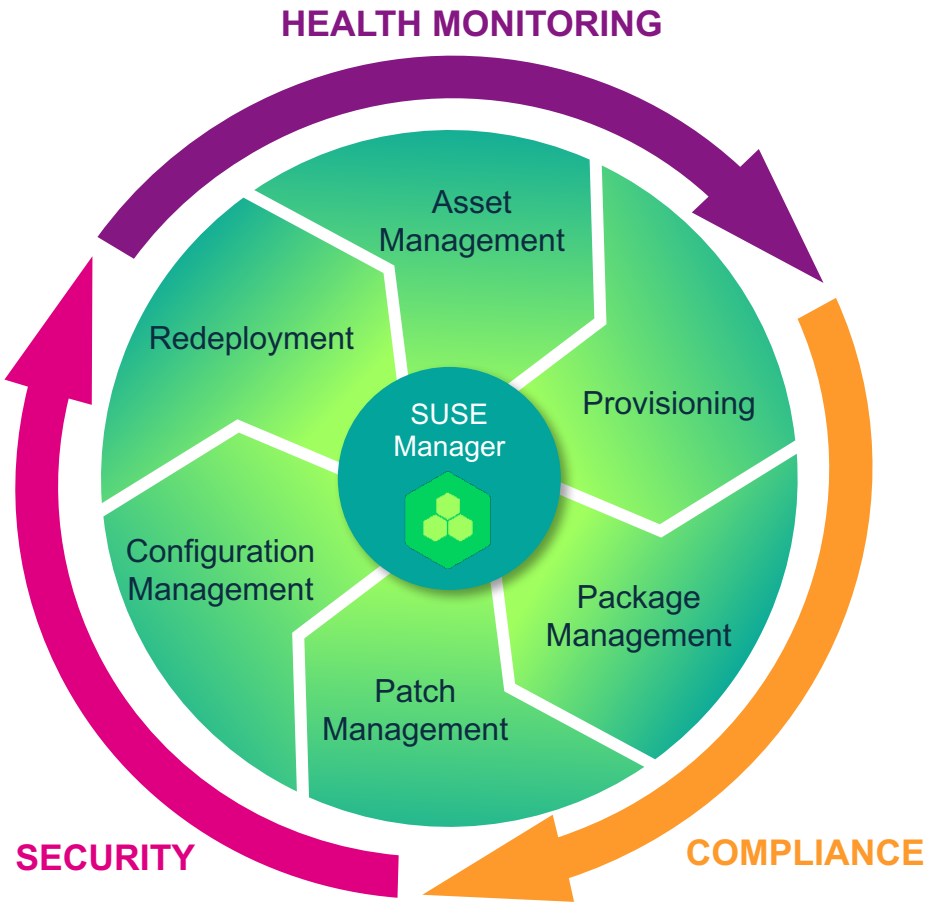
Patch & configuration management should be centralized

Multi-platform support

Many elements to Big Data environment -> Single source of support reduces contention

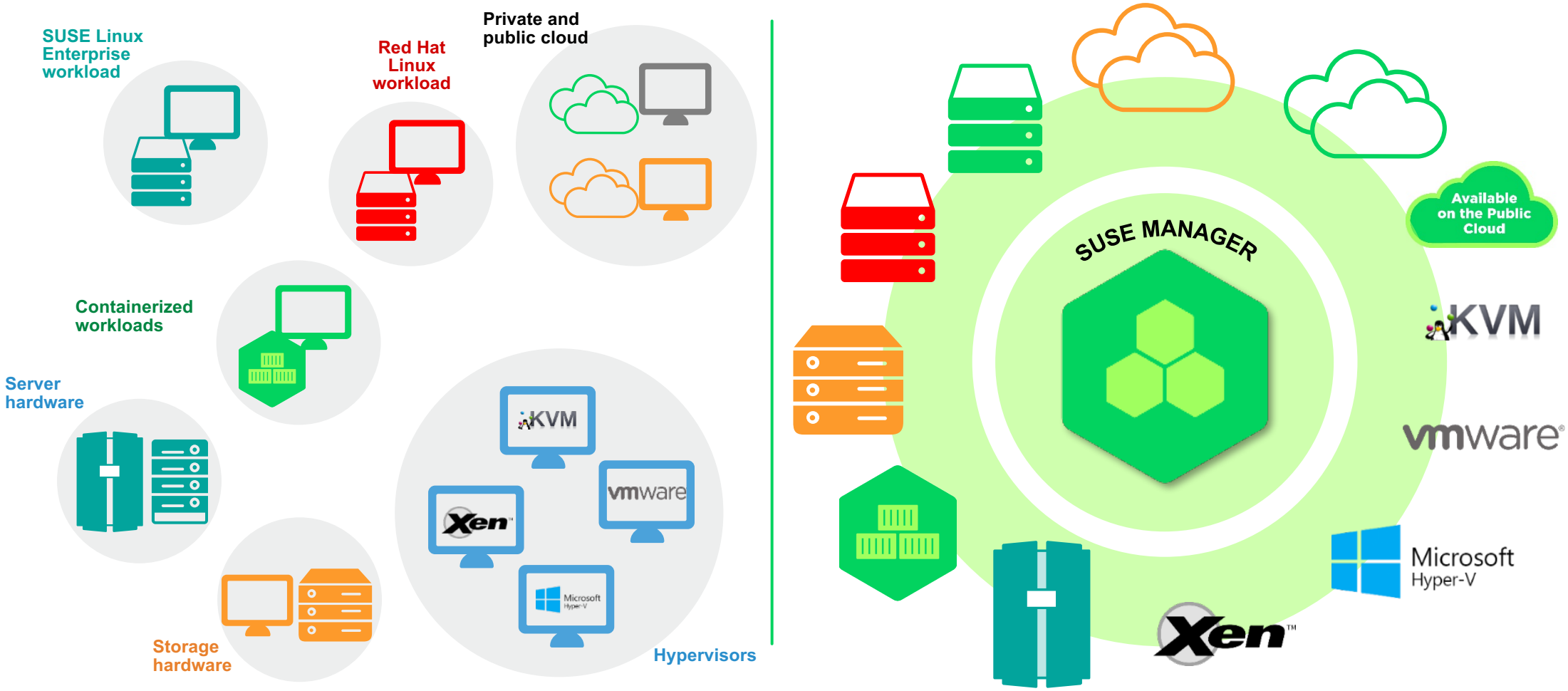
SUSE Manager

- Centralized management solution
- Patch & configure compute nodes
- Heterogeneous software environments
- Support multiple enterprise Linux distributions
- Hardware & virtualized platform agnostic
- VM’s & physical treated alike



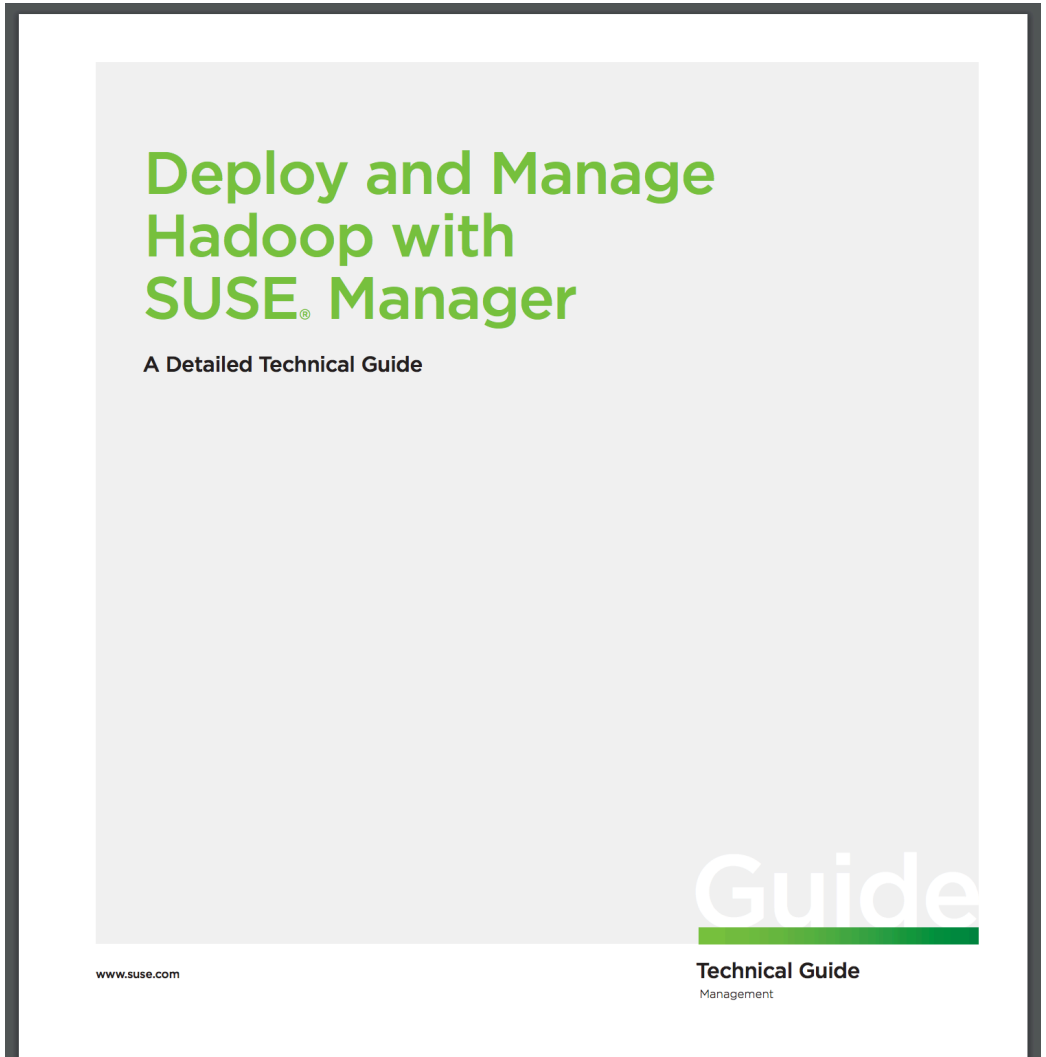
Reduce Complexity and Regain Control of Your IT

Using a single tool to manage Linux system configuration and compliance across a variety of hardware architectures, hypervisors and cloud platforms



Deploy and manage Hadoop with SUSE Manager

On physical, virtual or cloud environments



3.1 Compute for BigData

SUSE's history with OpenStack

- Founding member and platinum sponsor of OpenStack Foundation
- First to launch an "enterprise-grade" OpenStack distribution in 2012
- Focused on making SUSE OpenStack Cloud the best private cloud solution for enterprise business

Data Centre Cloud Vendor of the Year

Best Software defined Solution

Cloud Enabler of the Year

IaaS Solution of the Year

SUSE OpenStack Cloud

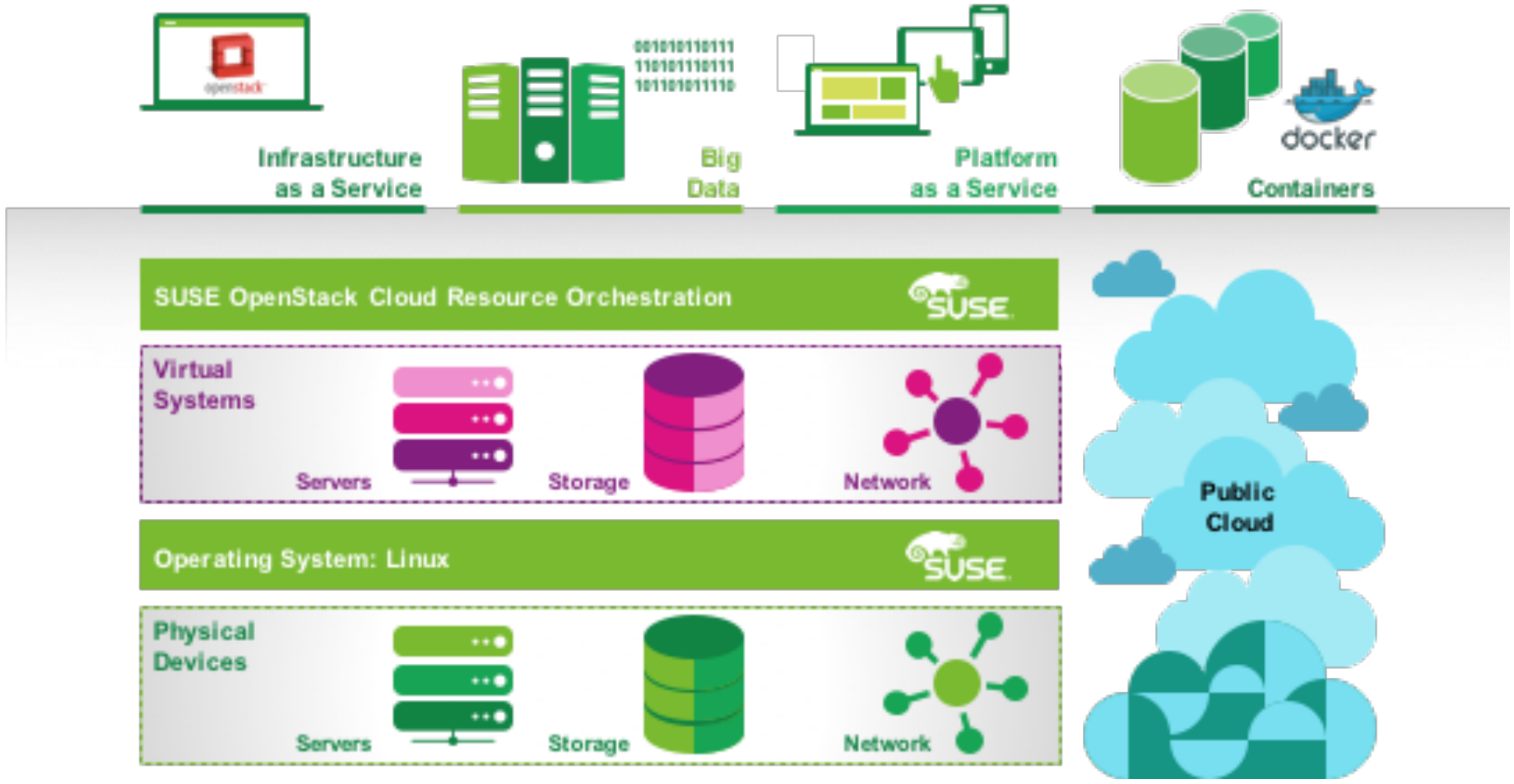
Answers all the data center challenges

SUSE OpenStack Cloud delivers:

- Flexibility to respond quickly & easily to new demands
- Increased agility, speed and efficiency
- An elastic platform for increased innovation
- Lower costs, faster ROI
- Greater control and security
- Self-service capabilities
- High quality services

SUSE OpenStack Cloud Architecture

Foundation for a Software Defined Data Center



Simplified BigData deployments with OpenStack Sahara

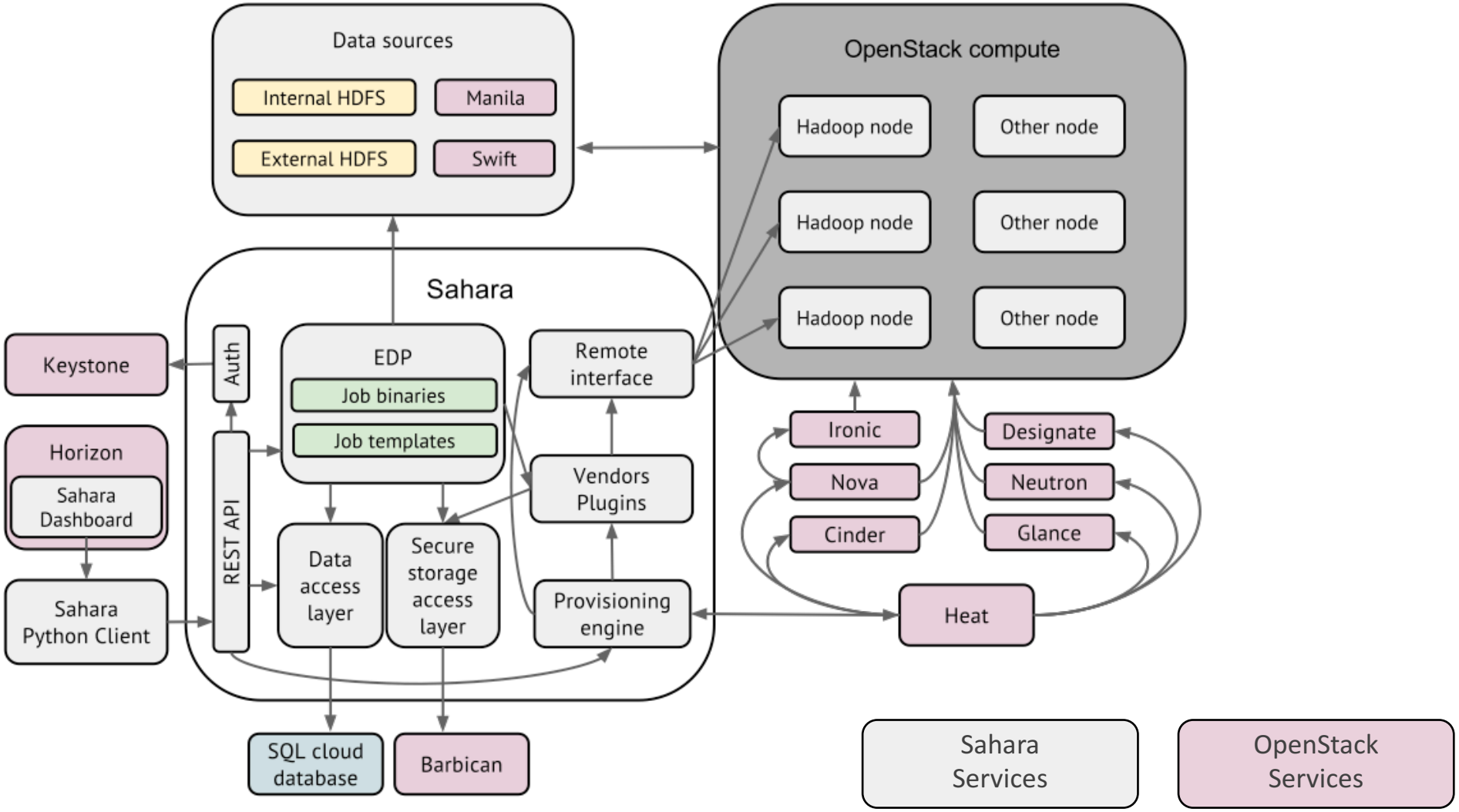
Use cases:

- Fast provisioning of data processing clusters on OpenStack for development and quality assurance(QA).
- Utilization of unused compute power from a general purpose OpenStack IaaS cloud.
- “Analytics as a Service” for ad-hoc or bursty analytic workloads

OpenStack Sahara – BigData frameworks

- **Multiple Hadoop vendor distributions.**
- **Apache Spark and Storm.**
- **Integration with vendor specific management tools, such as Apache Ambari and Cloudera Management Console.**
- **Pluggable system of Hadoop installation engines. Included plugins:**
  - Vanilla Plugin - deploys Vanilla Apache Hadoop
  - Ambari Plugin - deploys Hortonworks Data Platform
  - Spark Plugin - deploys Apache Spark with Cloudera HDFS
  - MapR Distribution Plugin - deploys MapR plugin with MapR File System
  - Cloudera Plugin - deploys Cloudera Hadoop

OpenStack Sahara architecture



3.2 Containers for BigData

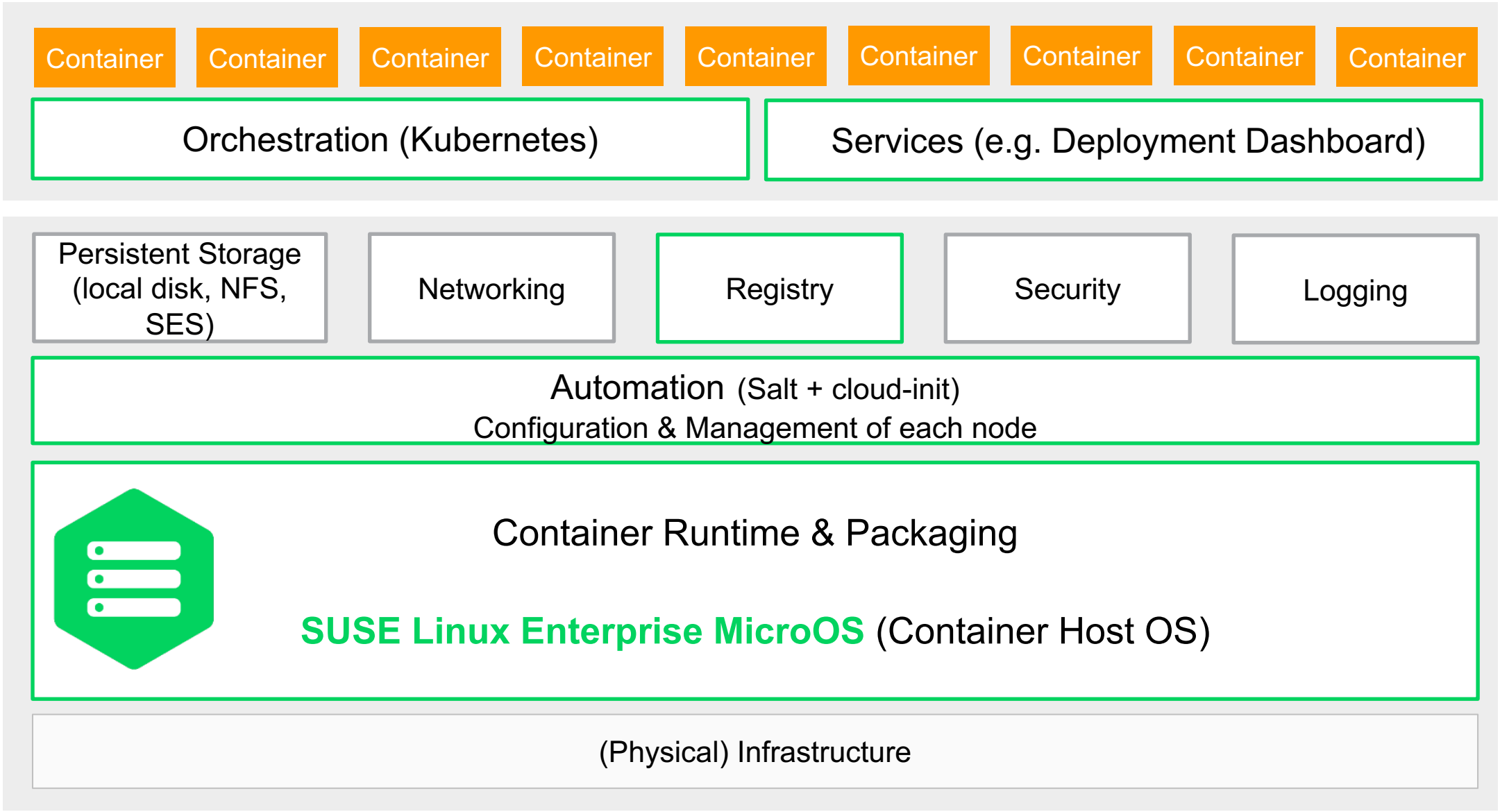
What is SUSE Container as a Service Platform

SUSE Container as a Service Platform is an infrastructure platform **for containers** that allows customers to **provision, manage, and scale container-based applications.**

Goal:
Deliver a container application development and hosting platform that **automates** the tedious **management tasks** allowing customers to **focus** on **development of applications** and meet their business goals faster.

SUSE CaaS Platform has 3 key components

SUSE CaaS Platform



Orchestration with Kubernetes

- Complete solution for container based workloads
  Deploy, Scale, Manage
- Self-healing
- Avoid vendor-lock-in
- Dashboard + Command line capabilities

Open Source Project Kubernetes

Kubernetes and BigData

- Additional resources optimization
- Declarative & scalable deployments
- Unique features: StatefulSets, (Anti)Affinity, Volume management, …
- Hundreds of reference deployments and architectures covering: Hadoop, Spark, Kafka, ZooKeeper, ELK, SAP Vora, …

3.3 Storage for BigData

SUSE Enterprise Storage

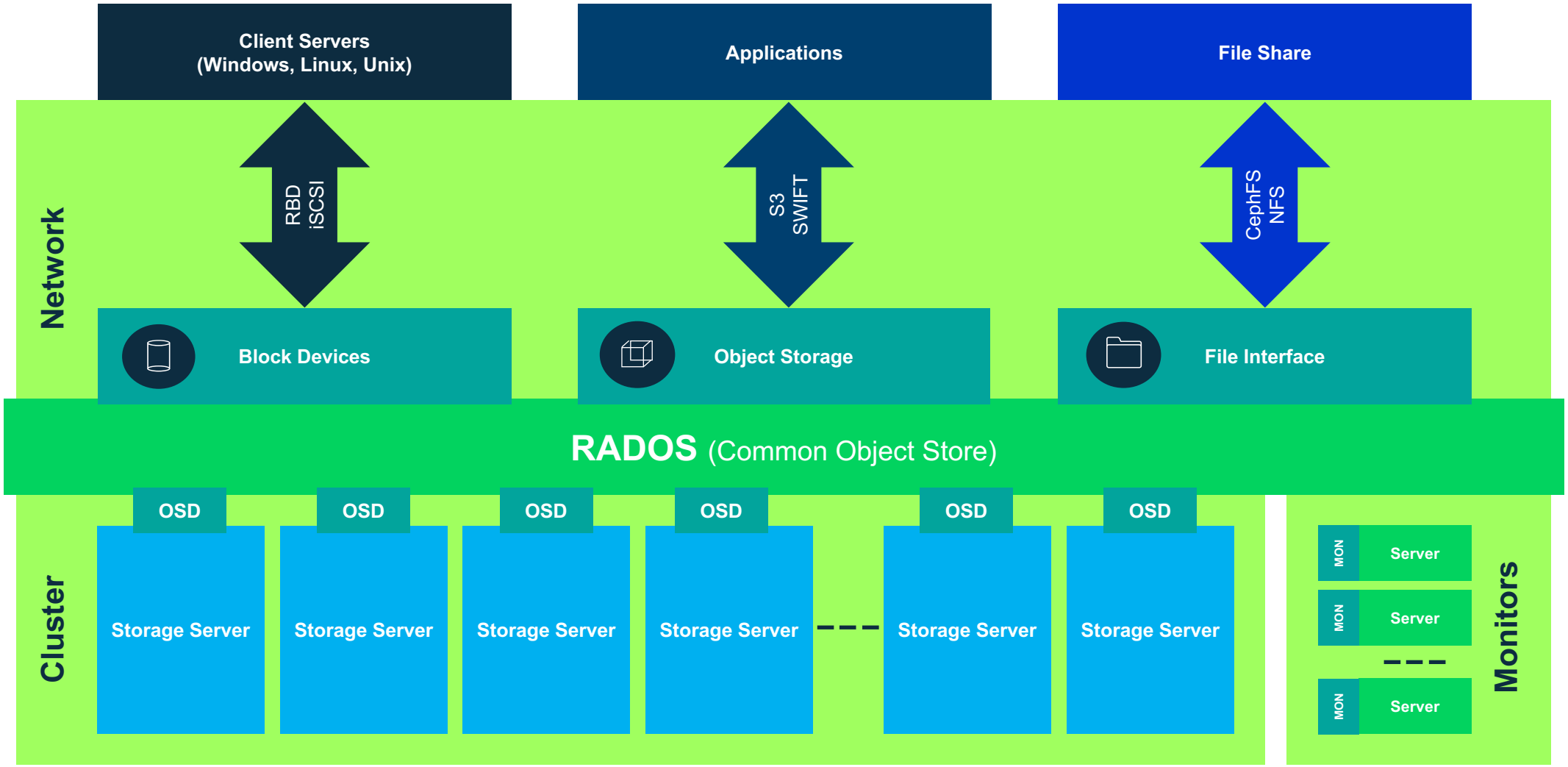
An intelligent software-defined storage management solution, powered by **Ceph** technology, that enables IT to transform their enterprise storage infrastructure to deliver highly scalable and resilient storage that is cost-efficient and able to seamlessly adapt to changing business and data demands

SUSE Enterprise Storage

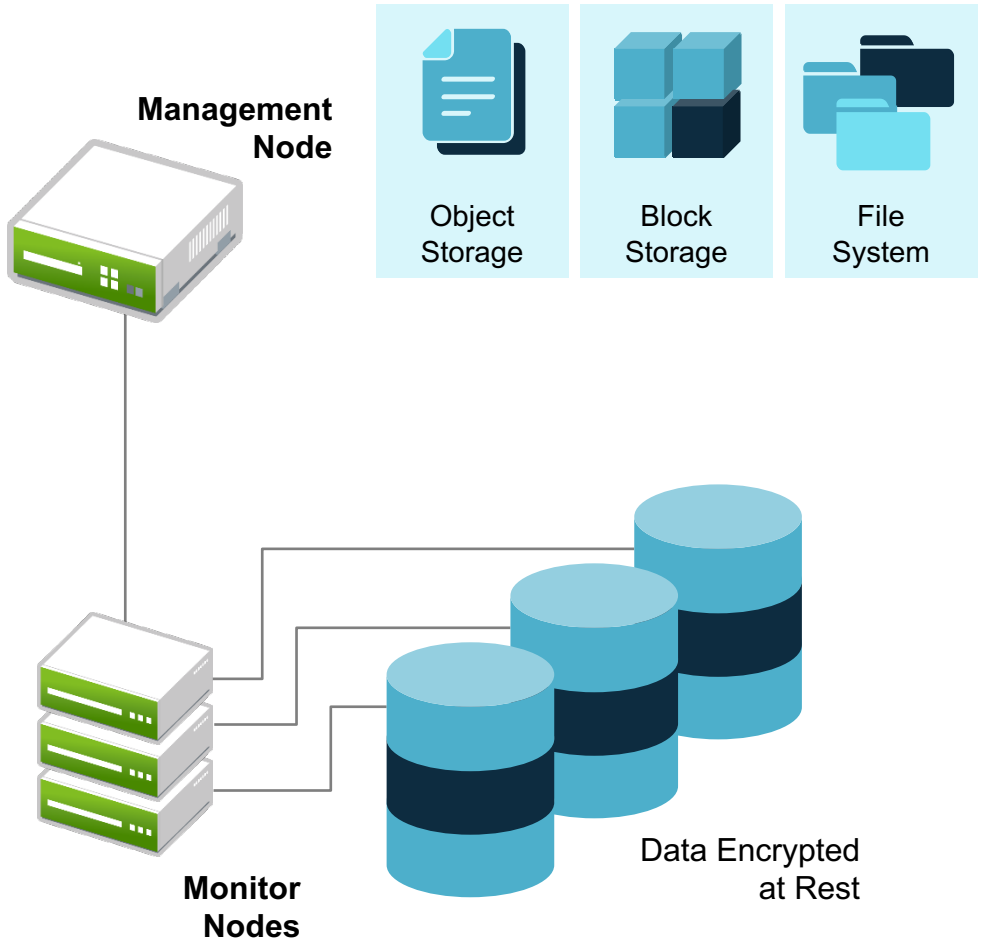
Seamlessly adapt to changing business and data demands

- Reduces IT capital and operational costs with an intelligent software-defined storage management solution
- Delivers resilient storage environment
- Unlimited scalability
- Consolidates block, object and file storage requirements
- Utilizes commodity off the shelf hardware
- Open
- Best solution of OpenStack, Large Data and bulk storage requirements

Open Source Ceph as the Base - Architecture



Truly Unified – Supports Object, Block and File System Storage in the Same Cluster

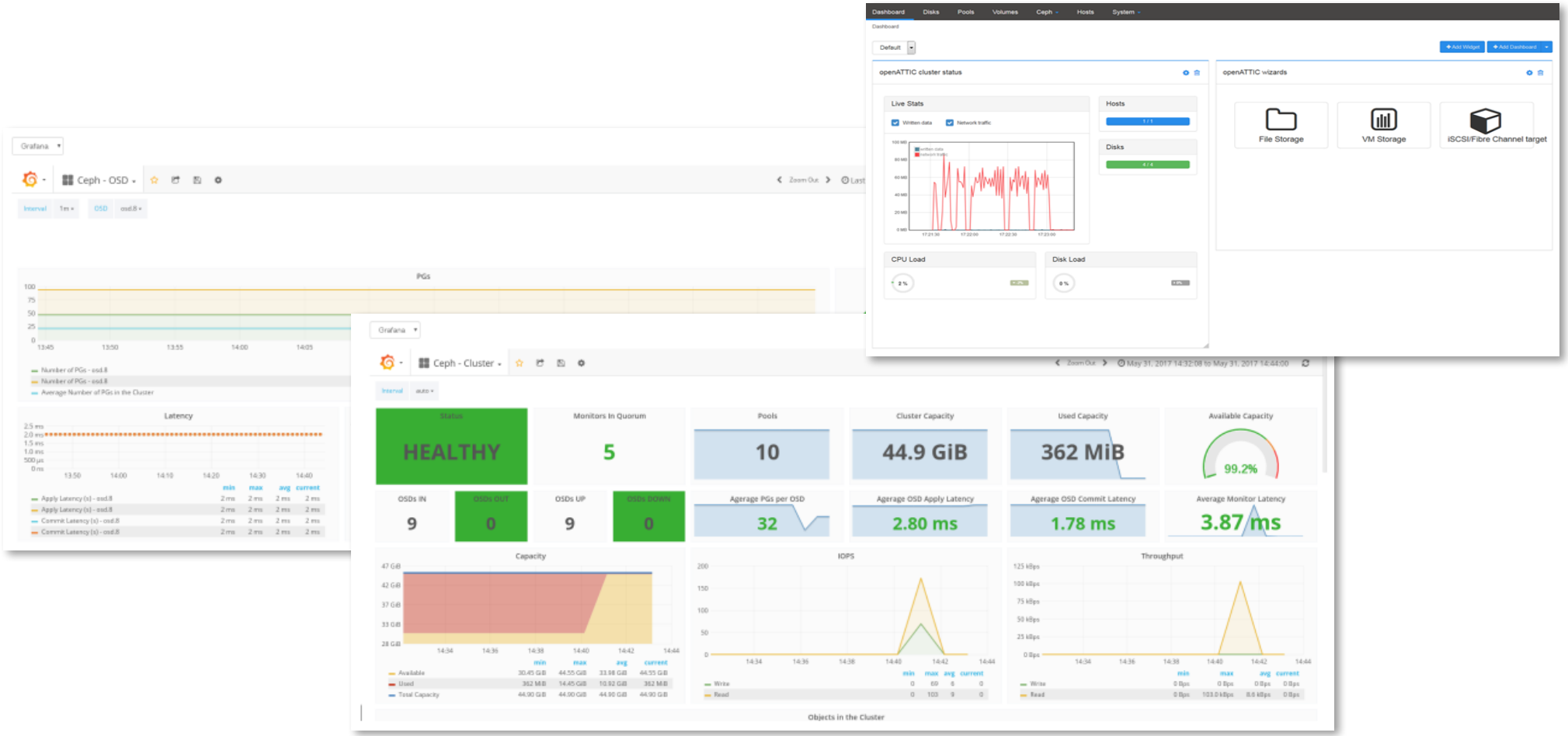


Automated management

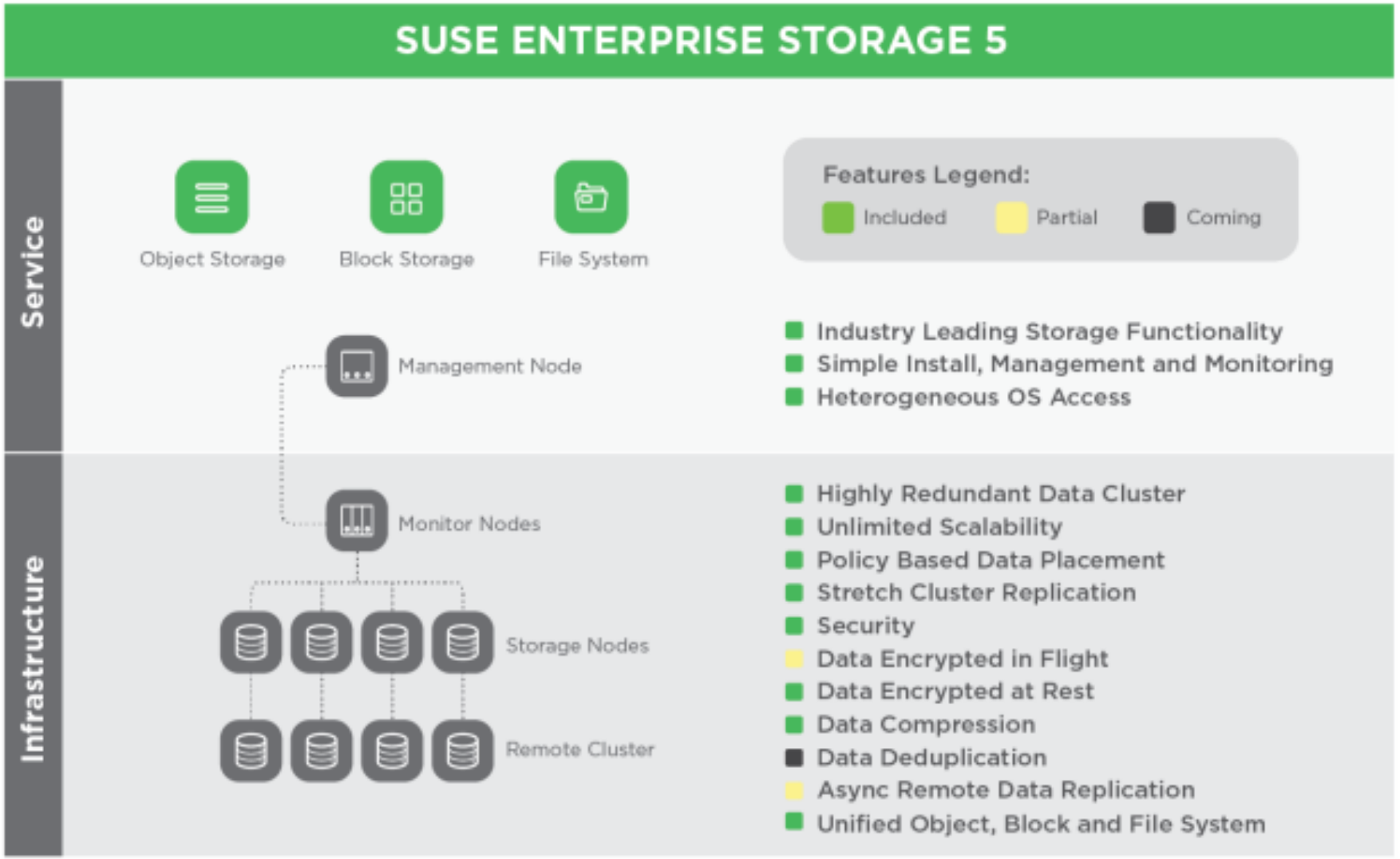
Heterogeneous OS Access

Major new features and benefits with SUSE Enterprise Storage 5

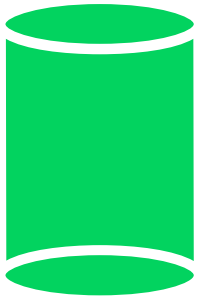
- Advanced graphical user interface for simplified management and improved cost efficiency, using the openATTIC open source storage management system.



SUSE Enterprise Storage 5 – Major Feature Summary

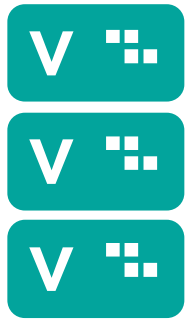


Perfect fit for Large Data Use Cases



Bulk Storage

Video Surveillance



Virtual Machine Storage

Data Archive

File System Use Cases

COMMVAULT VERITAS VEEAM Hewlett Packard Enterprise

Backup
- Solution supported with key backup ISVs (Commvault, Veritas, Veeam, HPE Data Protector)
- Ideal for low RTO & RPO backups*

Archiving
- Supported for healthcare archives, image and video archives, HPC Aachives
- Support ror many open source tools and commercial ISVs (Commvault, Veeam)
- Some customers have to deploy their own home-grown archival solutions

Analytics
- Certified configuration of SUSE Enterprise Storage is available for SAP Hana
- Storage for Hadoop supported through OEM partners

Commercial HPC
- Customers deploying SUSE Enterprise Storage in home directories of HPC clusters with high throughput requirements
- Not supported for low latency applications currently

*SUSE Supports these ISVs with Block and file (thru CephFS). Both Block and File show similar performance.

Home Directories
- Large home directories supported in various use cases
- Examples: HPC storage, application binaries and data (video and images, medical data)
- Linux home directories

Object Storage Use Cases

VERITAS

COMMVAULT

Backup

- Solution supported with key backup ISVs (Commvault and Veritas)
- Ideal for high-throughput requirements through S3

Archiving

- Supported and deployed for healthcare archives, image and video archives, HPC archives
- Supported with commercial ISVs (Commvault, etc.) and open source tools (DMF)

Analytics

- Supported through S3A only for Hadoop and Spark
- Repositories for unstructured data (IoT data, log repositories, social data, etc.)

Cloud Storage

- Supported for CSP use cases (modern email, S3 Storage, File Sync and Share etc.)
- Supported with Openstack.* Preferred storage Solution for SUSE Openstack Cloud
- Use cases include image repos. and data storage

*SUSE, with the acquisition of HPE Helion, has over 23% of the marketshare in OpenStack Cloud.

Content Storage and Distribution

- Supported for use cases of video surveillance, repositories and distribution
- Support for multiple protocols (Swift, S3) and 1:1 and 1:Many distribution

Block Use Cases

COMMVAULT VERITAS VEEAM Hewlett Packard Enterprise

Backup

- Solution supported with key backup ISVs (Commvault, Veritas, Veeam, HPE Data Protector, EMC Networker)
- Ideal for low RTO/RPO use cases

Compliance Archives (with iTernity)

- Archiving supported with iCAS for compliance
- Fits requirements of WORM storage

Analytics

- Certified configuration of SUSE Enterprise Storage is available for SAP Hana

Cloud Storage

- Supported with Openstack.* Preferred storage solution for SUSE Openstack Cloud
- Use cases are primary storage (with Cinder)

*SUSE, with the acquisition of HPE Helion, has over 23% of the marketshare in OpenStack Cloud.

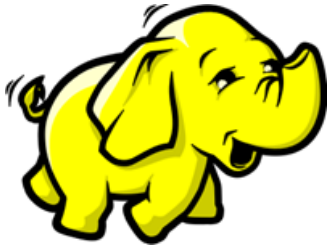
VM

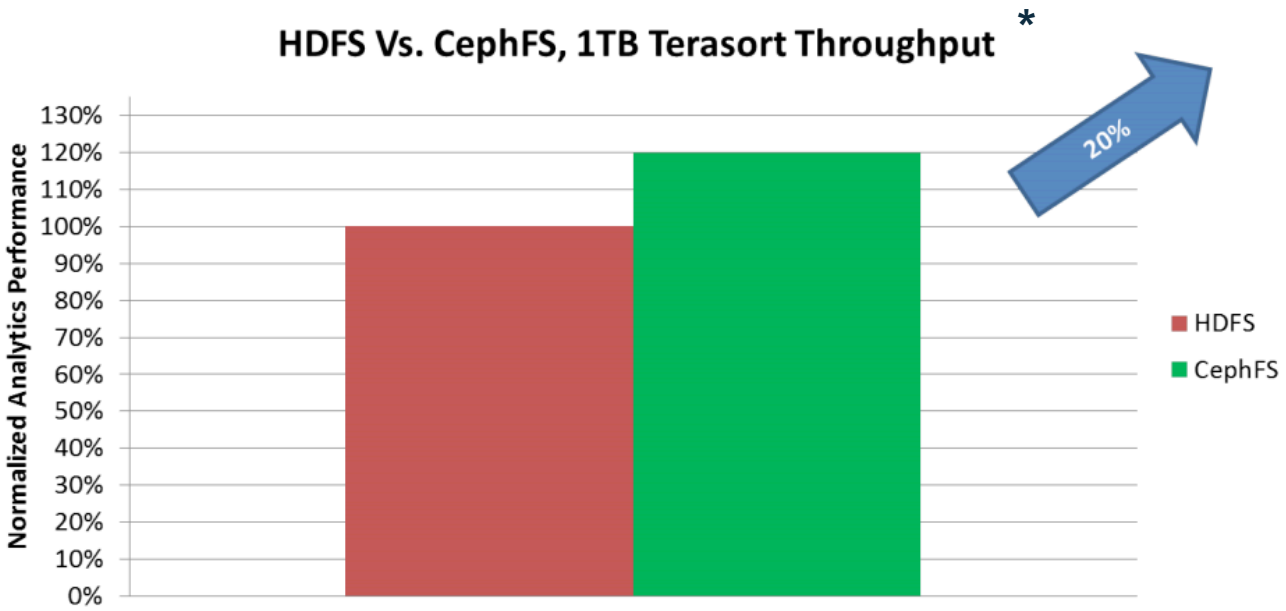
VM Storage

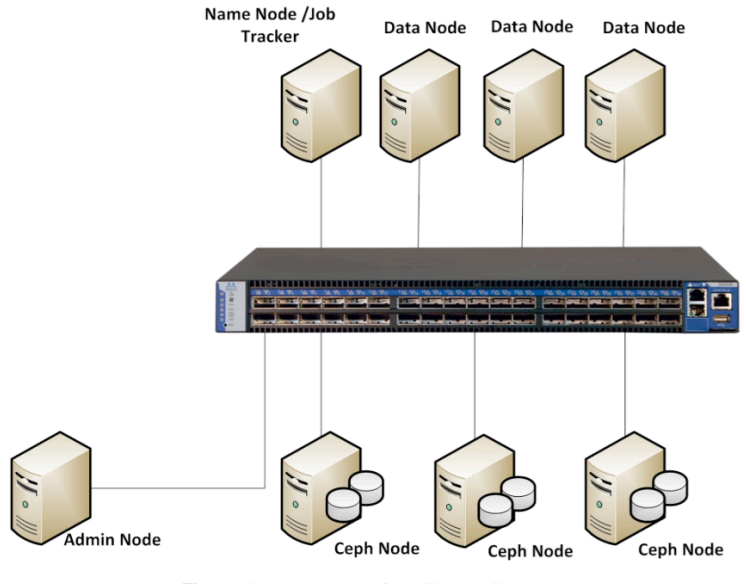
- RBD for KVM
- iSCSI for VMware and Hyper-V

Ceph as HDFS replacement

- Increase Hadoop Cluster Performance
- Scale Compute and Storage solutions in Efficient Ways
- Single Point of Failure event in Hadoop Architecture







Source: Mellanox http://www.mellanox.com/related-docs/whitepapers/wp_hadoop_on_cephfs.pdf

4. BigData ISVs / Partners

Big Data partners

Certified and recommend SUSE Linux Enterprise Server as operating system for their solutions.

- **SAP Hana & Vora**
- **Vertica**
- **Teradata**
- **Cloudera**
- **Hortonworks**
- **MapR**
- **InterSystems**
- **MongoDB**
- **Revolution Analytics**

5. BigData IHV

Reference architectures

SAP Global Partner

HITACHI
Inspire the Next

SAP HANA Scale-Up High Availability with HANA System Replication a
SUSE High Availabilit

TECHNICAL PAPER

Reference Architectur

By Milind Pathak

January 2016

ARCHITECTURE BRIEF
BIG DATA

Lenovo

Lenovo Big Data Reference Architecture for Cloudera Distribution for Hadoop

Easy to implement hardware, software and services for big data analytics architecture

Architecture Highlights

- Provides guidance for deploying Lenovo systems with a Cloudera distribution for Hadoop to coordinate the processing of the data across a massively parallel environment.
- Includes the latest data center equipment available such as the Lenovo x3650 M5 and x3550 M5 and Lenovo RackSwitch Ethernet switches and Lenovo XClarity.
- Supports entry through high-end configurations and the ability to easily scale as the use of big data grows

Big data is more than a challenge. It is an opportunity to find new insights in data to make your business more agile and to answer questions that were previously beyond reach. Today, Cloudera uses the latest big data technologies such as the massive map-reduce scale-out capabilities of Hadoop to open the door to a world of possibilities.

This Lenovo Big Data RA for Cloudera Distribution for Hadoop is certified by Cloudera and provides a thoroughly tested and integrated solution that combines the benefits of leading-edge technologies with mature, enterprise-ready features. Starting with a preconfigured hardware platform that is Cloudera-certified helps your team to be up and running analytics quickly.

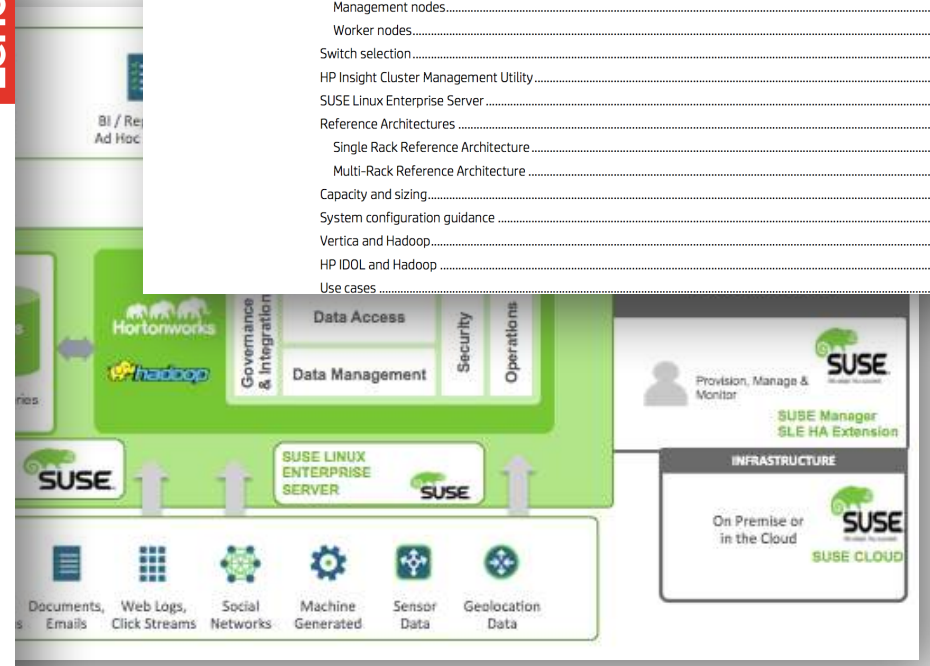
Cloudera allows organizations to run large-scale, distributed analytics jobs on clusters of cost-effective server hardware. This infrastructure can be leveraged to tackle very large data sets by breaking up the data into "chunks" and coordinating the processing of the data across a massively parallel environment.

Technical white paper

HP Reference Architecture for Hortonworks Data Platform 2.1 on ProLiant DL Servers – SUSE Linux Enterprise Server

HP Converged Infrastructure with Hortonworks Data Platform 2.1 for Apache Hadoop

Table of contents



Summary

SUSE & BigData: Summary

- SUSE addresses Big Data requirements
- SLES is a flexible, available platform
- High performance, large data access capabilities
- #1-rated support
- SUSE Manager for multi-system management
- SUSE Cloud for distributed deployment
- SUSE Kubernetes for bleeding edge deployment architectures
- SUSE Storage for resilient and unlimited scale data management
- Big Data ISVs support SUSE as underlying OS including Cloudera, Hortonworks, Intel, ...
- Solution partners have chosen SUSE for integrated Big Data solutions including SAP, Teradata, and Intersystems
- Part of Microfocus group with top level solutions like Vertica


SUSE
We adapt. You succeed.