

# Practical Architecture for SAP Data Hub on Premise

Nathan Saunders, SAP North America Alliances

Jose Betancourt, SUSE Global Alliances Director

Lenovo

# Data is kept in Silos across the Enterprise

Plain Files



IoT



Mobile



Cloud Storage



Databases



VALUE



Enterprise Applications



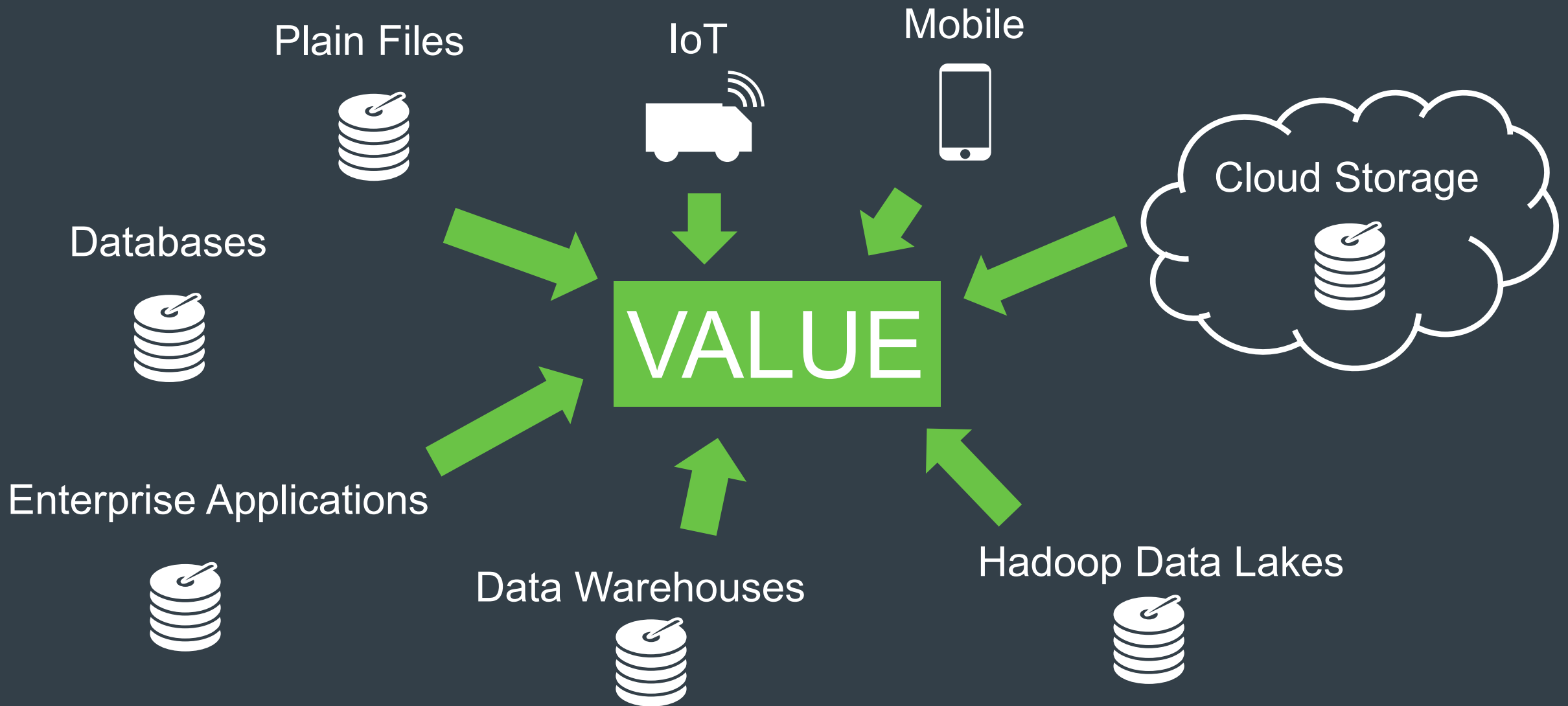
Data Warehouses



Hadoop Data Lakes



# Unlock the Value of your Data



# Lenovo Solution for SAP Data Hub

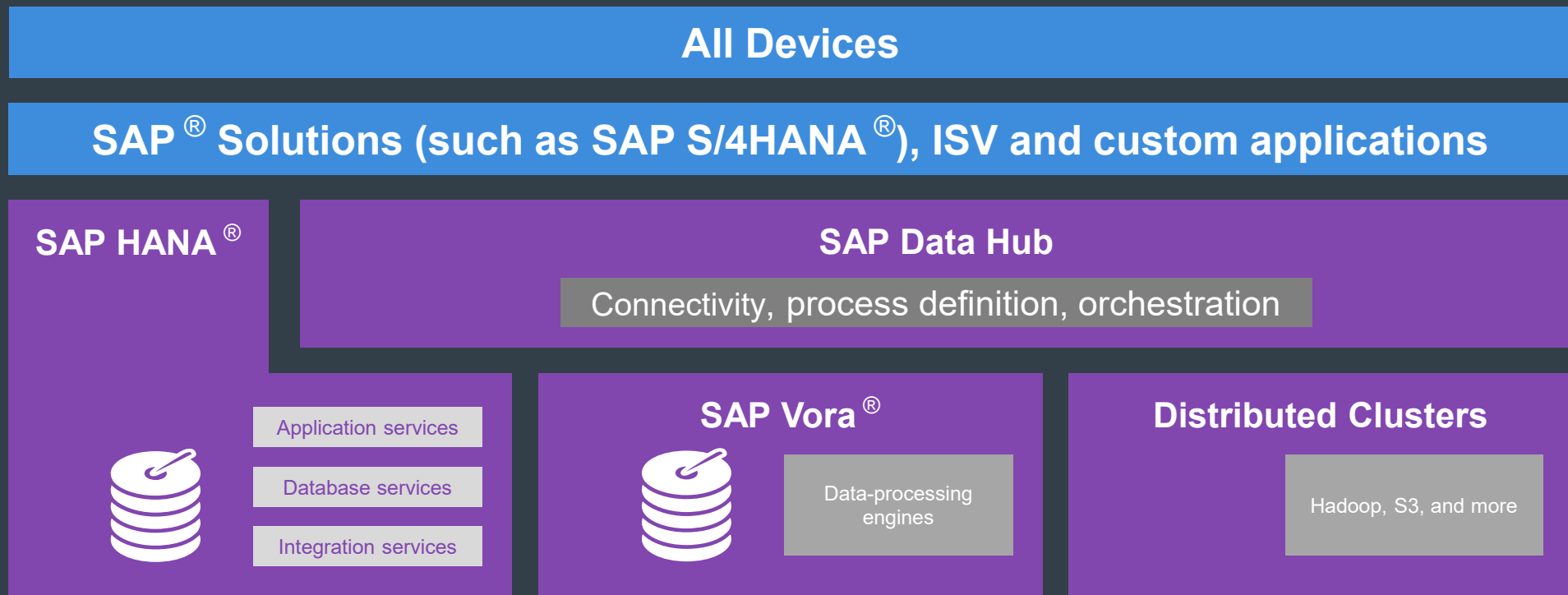
Simplified Deployment, Accelerated Implementation

Universal view of enterprise and Big Data from an **Engineered Solution** that provides a **flexible, scalable** architecture, and is **easy** to deploy.

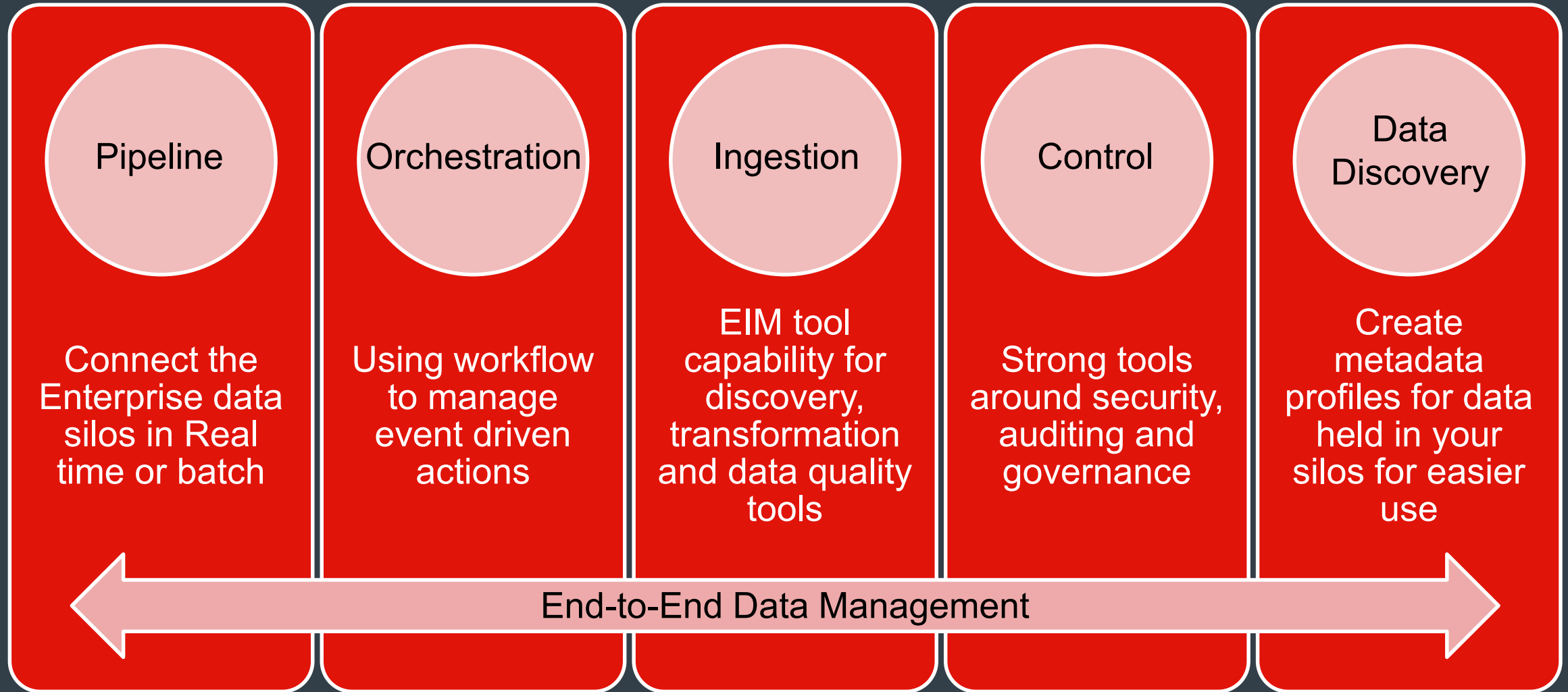
!

# An SAP Approach to Enterprise Data Management

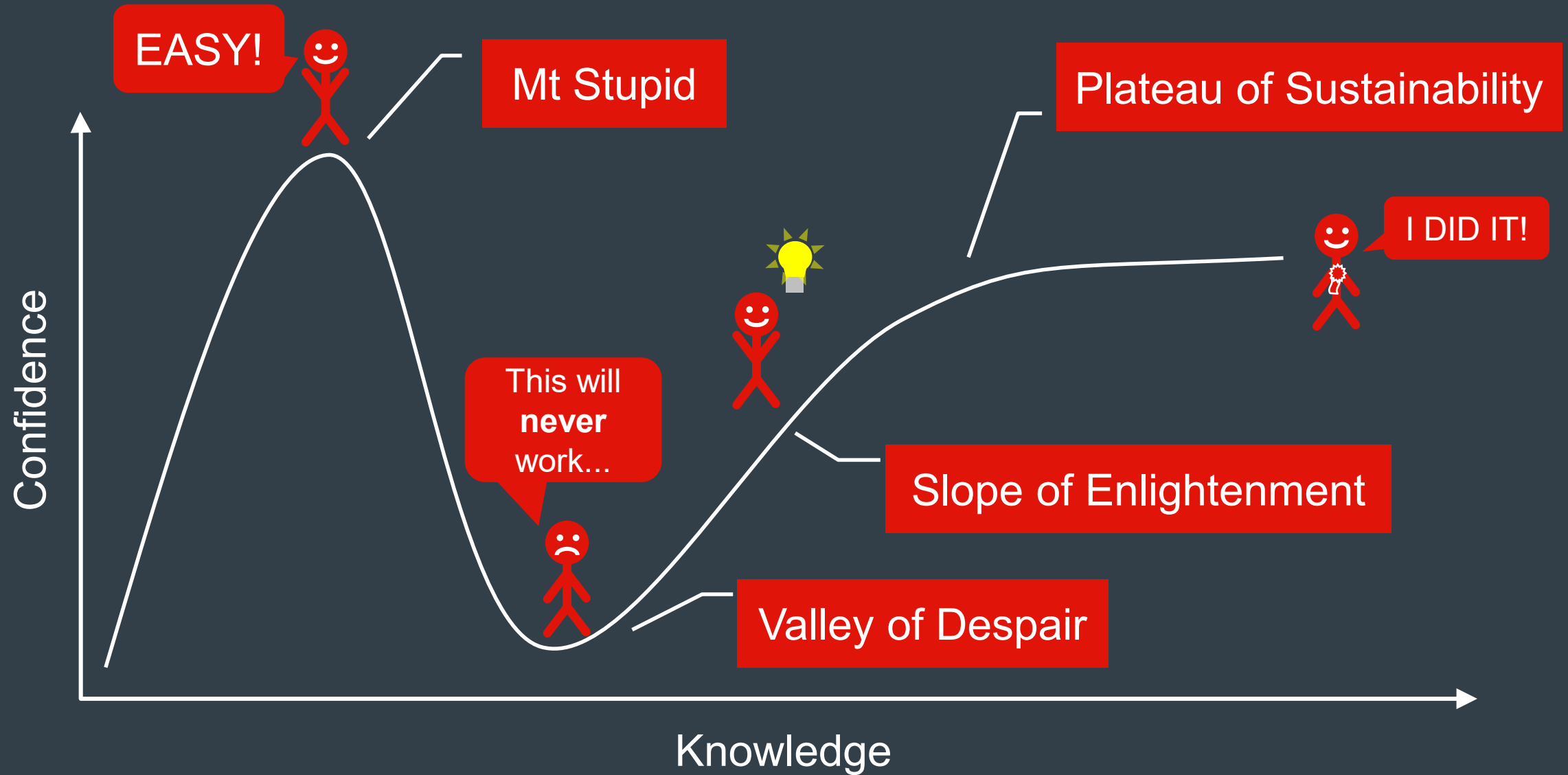
## Application-centric view



# What does SAP Data Hub actually do?




# Why a Reference Architecture?



# What do we want?

Scalability 

Performance 

Compact Footprint 

Flexibility 

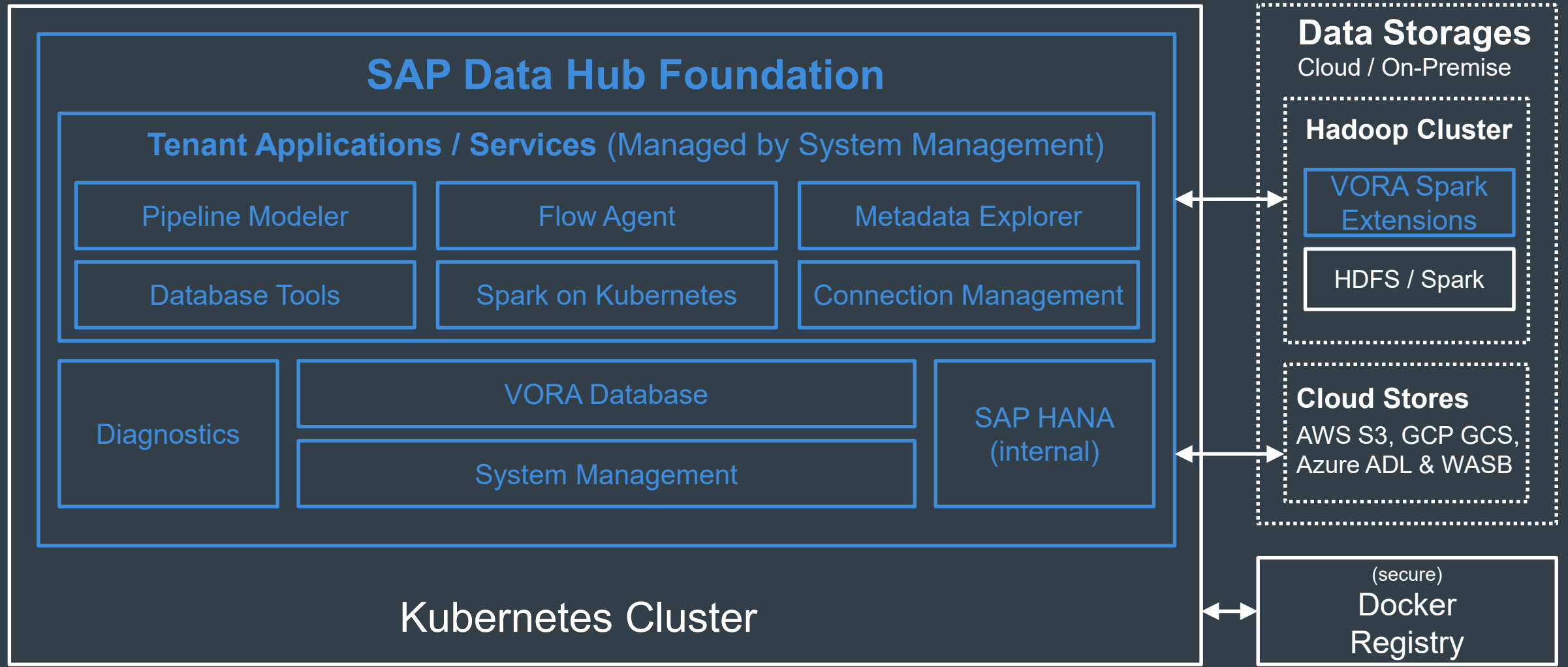
Ease of Deployment 

Security 

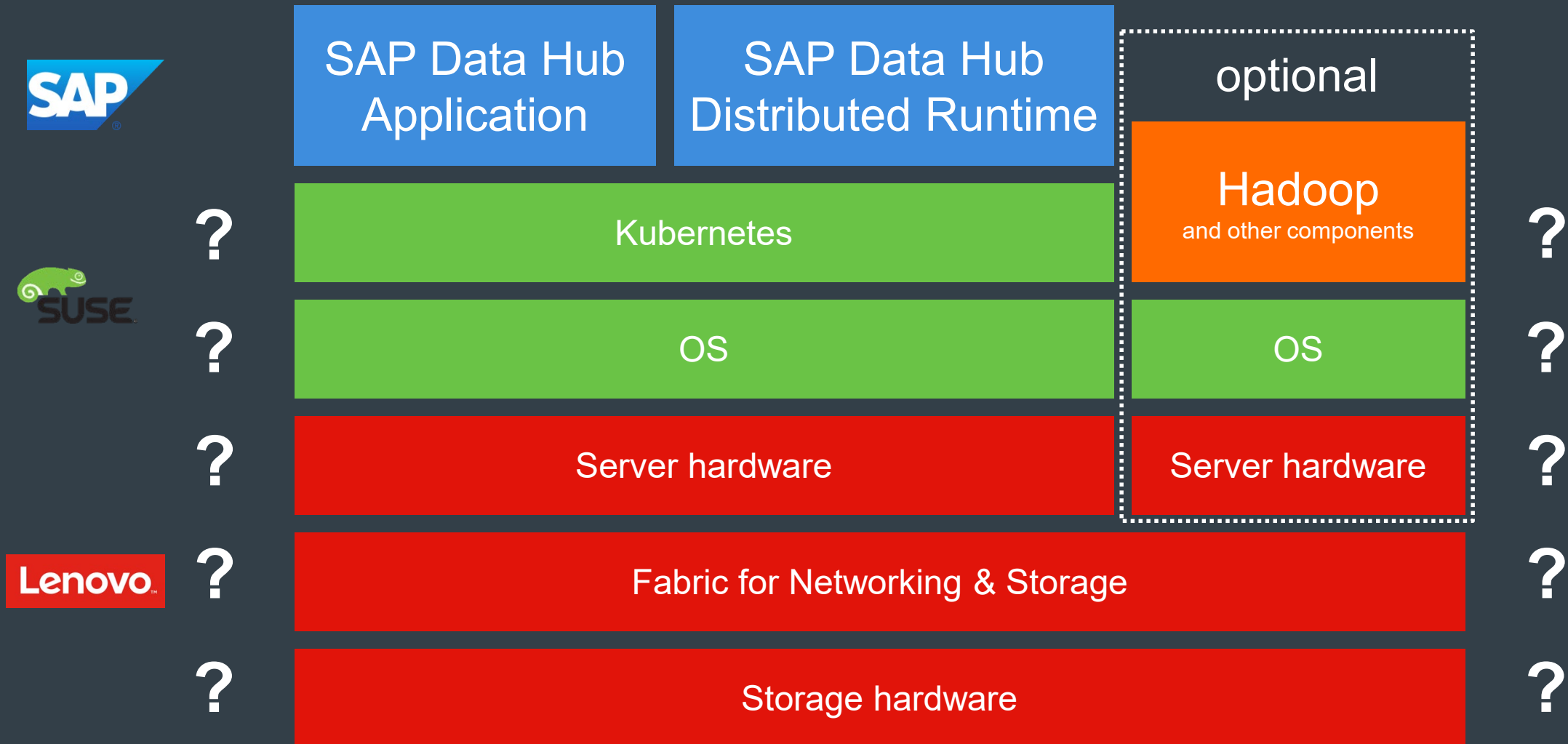
Fault Tolerance 



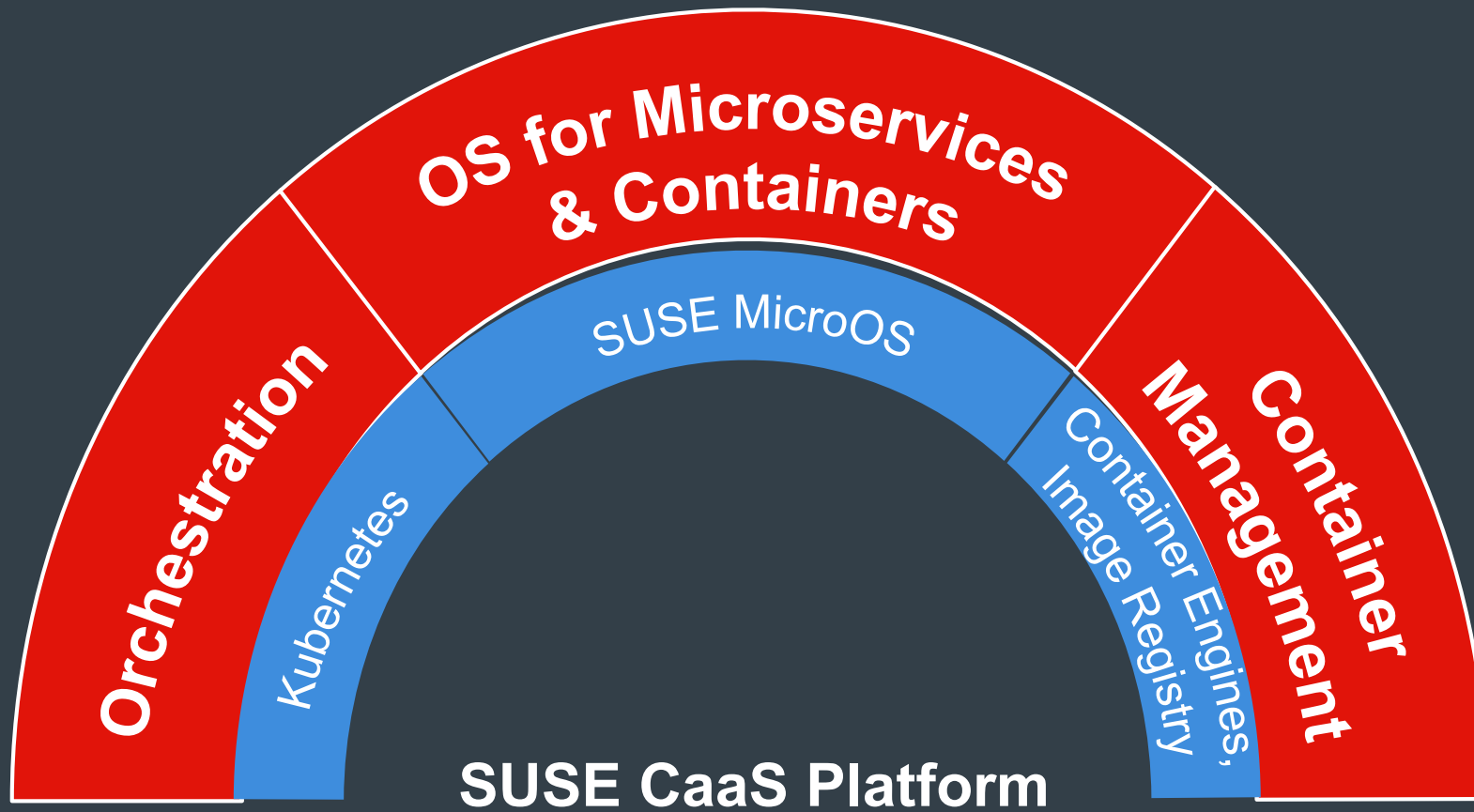
# SAP Data Hub Architecture



# SAP Data Hub implementation components



# Three Key Technologies, One Enterprise Platform



SUSE CaaS Platform is an enterprise class **container management solution** that enables IT and DevOps professionals to more **easily deploy, manage, and scale** container-based applications and services.

# Kubernetes Infrastructure

- Scalability: just add more nodes
- Flexibility: Come on, it's Kubernetes 😊
- Performance: 2 Intel Xeon SP CPUs per node
- Fault tolerance: Survives the loss nodes
- Security:
  - Role-Based Authentication for everything
  - Secure Docker Registry (Portus) for SAP Data Hub images
- Compact footprint: 1U nodes
- Ease of Deployment: detailed, easy to follow setup-instructions
- Enterprise-level support from SUSE and Lenovo

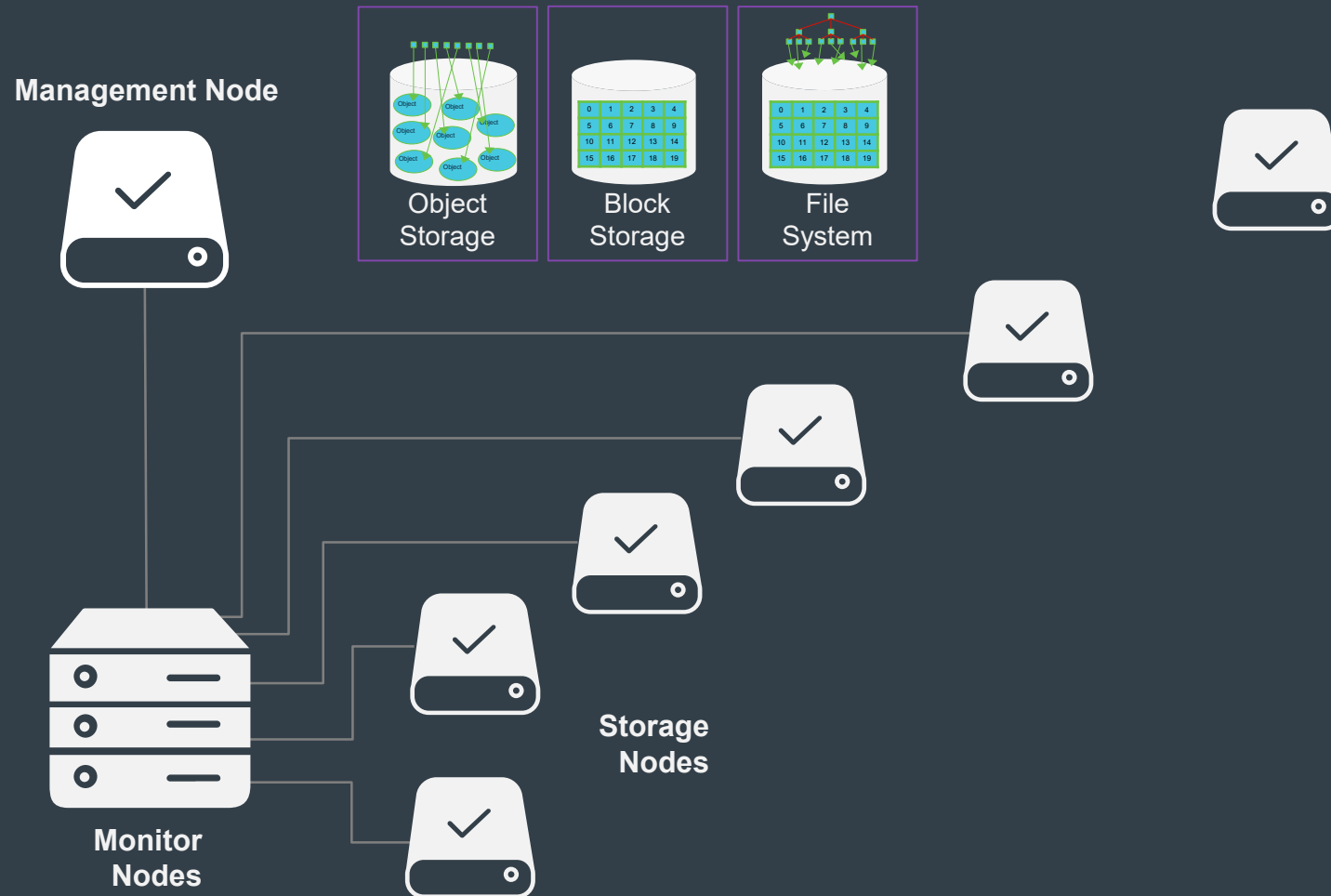
Powered by Intel® Xeon®  
processor Scalable family



Lenovo

# SUSE Enterprise Storage

Unlimited Scalability with self-managing Technology



- Highly Scalable
- Reduces Cost
- Seamlessly adapts to business needs
- Interfaces:
  - Block: iSCSI, RBD
  - Object: S3-API
  - File: CephFS, NFS
  - Integrates with CaaSP
- High Performance\*

\* Lenovo DSS-C offering certified as SAP HANA Enterprise Storage

# Storage

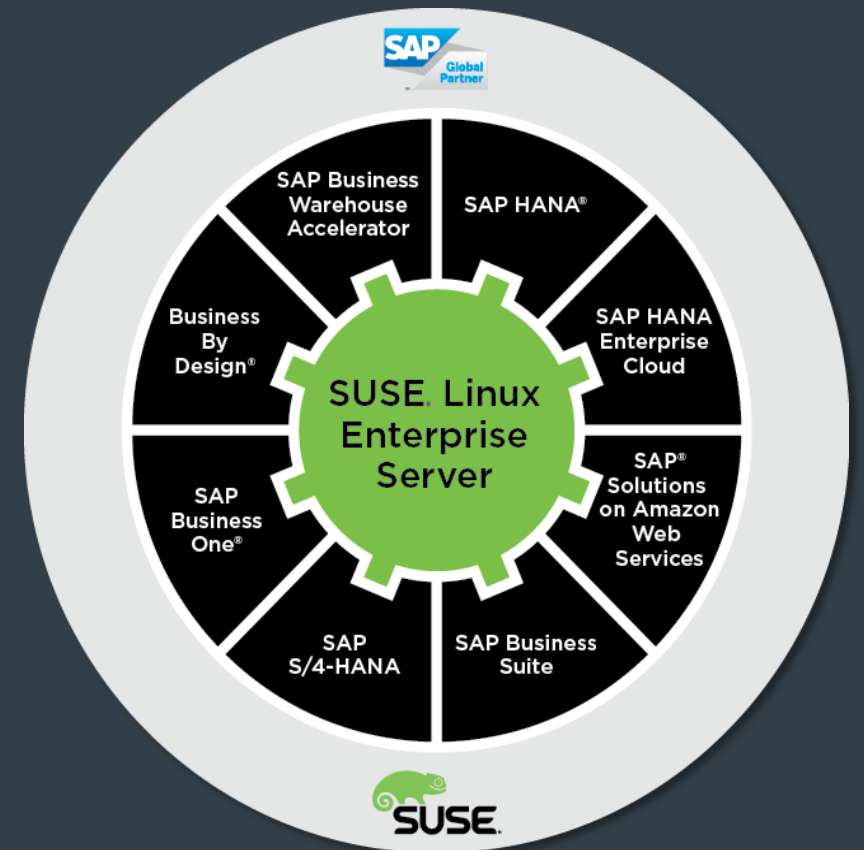
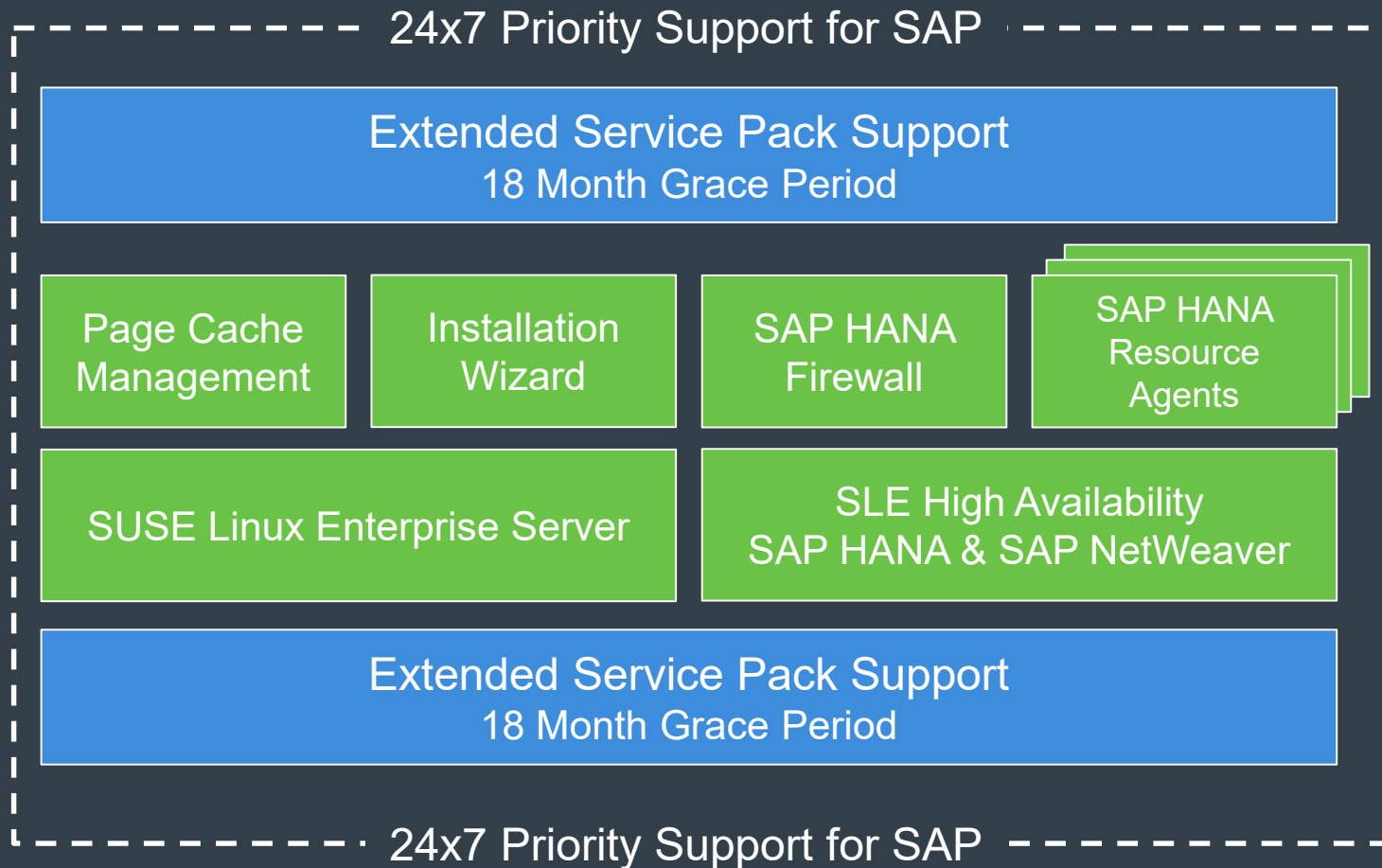
- Scalability: Extremely scalable
- Flexibility: Provides multiple ways of using the storage
  - dynamically provisioned persistent volumes for Kubernetes,
  - RBD volumes for everything else (Hadoop, Registry, ...)
  - S3 Object Storage
- Performance: entirely SSD based
- Fault tolerance: Survives the loss of drives, nodes
- Security: Authentication on storage-level
- Compact footprint: only 4 nodes = 8U Rackspace (but extensible if needed)
- Ease of Deployment: detailed, easy to follow setup-instructions
- Enterprise-level support from SUSE and Lenovo

Powered by Intel® Xeon®  
processor Scalable family



Lenovo

# SUSE Linux Enterprise for SAP Applications



# Hortonworks Data Platform



## Open

- Hortonworks is committed to an open approach to software development that spurs innovation
- HDP enables enterprises to deploy, integrate and work with unprecedented volumes of structured and unstructured data
- HDP delivers enterprise-grade software that fosters innovation and prevents vendor lock-in



## Interoperable

- HDP is interoperable with a broad ecosystem of data center and cloud providers
- HDP minimizes expense and effort required to connect customer's IT infrastructure with HDP's data and processing capabilities
- With HDP, customers can preserve their investment in existing IT architecture as they adopt Hadoop



## Central

- HDP is based on a centralized architecture supported by YARN that allocates resources among various applications
- YARN maximizes data ingestion by enabling enterprises to analyze data to support diverse use cases
- YARN coordinates cluster-wide services for operations, data governance and security

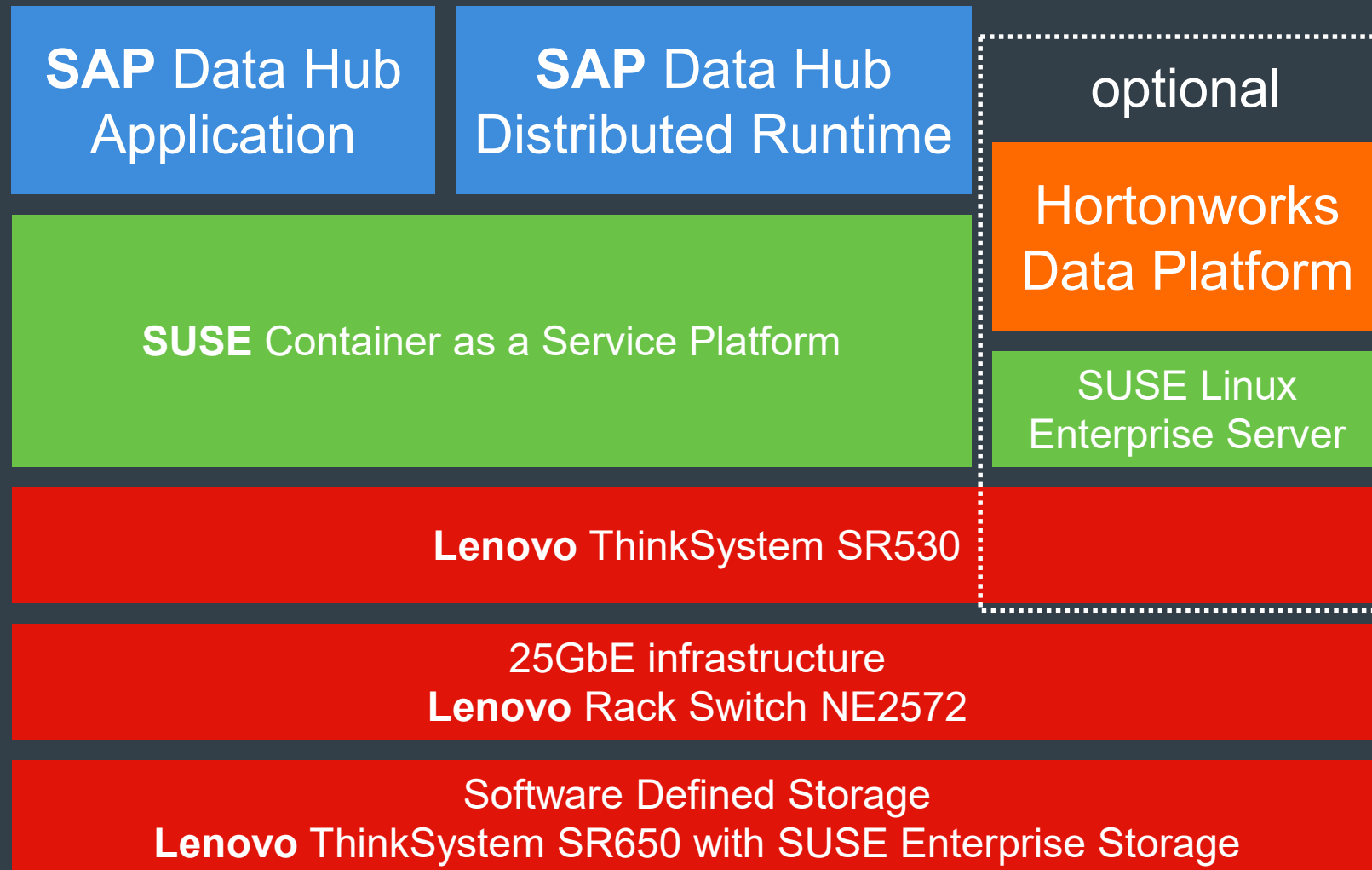


## Enterprise Ready

- HDP provides centralized Hadoop cluster management and monitoring
- With HDP, security and data governance is built into the platform
- HDP ensures that security is consistently administered across data access engines



# SAP Data Hub implementation components



# Hardware Building Blocks

## Storage Nodes



### Lenovo ThinkSystem SR650

- 2-socket, 2U Rack Server
- 2x 2.5" Intel S4500 SSDs for OS
- Up to 22x 2.5" Intel S4500 SSDs for Data
- 2 or 4 ports 25Gb Ethernet
- industry-leading reliability, management, and security

## Compute Nodes



### Lenovo ThinkSystem SR530

- 2-socket, 1U Rack Server
- 2x Intel Xeon SP CPUs
- 192GB / 384GB / 768GB RAM
- 2x 2.5" Intel S4500 SSDs for OS
- 2 ports 10Gb Ethernet
- industry-leading reliability, management, and security

## Fabric



### Lenovo ThinkSystem NE2572

- 10 Gb/25 Gb Ethernet connectivity
- 40 Gb/100 Gb Ethernet upstream links
- 100% line-rate performance
- Non-blocking architecture
- redundant hot-swap hardware components

Powered by Intel® Xeon®  
processor Scalable family



Lenovo

# Physical View

Powered by Intel® Xeon®  
processor Scalable family



## Compute Nodes

Lenovo ThinkSystem SR530  
1 or 2 CPUs, 192/384/768 GB RAM per Node  
Drives only for OS, Data on Storage  
SLES 12 SP3 / SUSE CaaSP 3

Redundant Fabric for Networking & Storage  
10GbE for Compute, 25GbE for Storage, 1GbE for Management

## Storage Nodes

Lenovo ThinkSystem SR650  
4+ 960GB SSD Drives per Node  
SUSE Enterprise Storage 5



## Hadoop Data Lake

### SAP Data Hub

3+ Kubernetes Worker  
1+ Kubernetes Master  
Docker Image Repository for SAP Data Hub  
CaaSP Cluster Management

Connectivity to Managed Systems  
or external Data Sources / Sinks, Up to 100GbE

## Storage for everything

Dynamically provisioned storage for  
SAP Data Hub  
Storage for Docker Registry  
Storage for Hadoop nodes  
S3 Object Storage

# Infrastructure Sizing for SAP Data Hub

- SUSE CaaS Platform nodes

- 1 Kubernetes master node (more for HA)
- 3 Kubernetes worker nodes (more depending on workload)
- 1 node for the private Registry (docker + Portus)
- 1 lifecycle management node (Velum)

Scalable:  
Just add more nodes

- SUSE Enterprise Storage nodes

- 4 nodes minimum
- 4 data drives per node minimum (up to 22)

Scalable:  
Just add more nodes

- Minimum Number of HDP nodes

- 3 nodes minimum

Scalable:  
Just add more nodes

Powered by Intel® Xeon®  
processor Scalable family



Lenovo

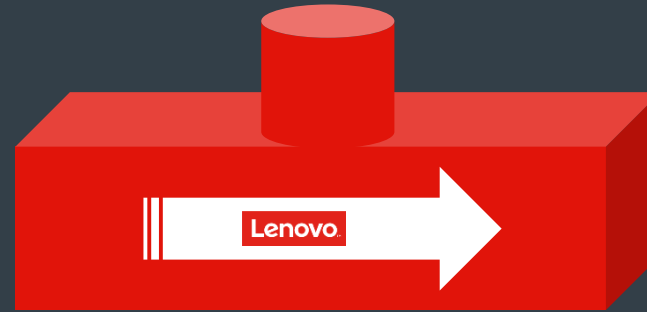
# Infrastructure Sizing for SAP Data Hub (simplified)

Pipeline Engine  
Vora Relational In-Memory Engine  
Vora Disk Engine  
Vora Time Series Engine  
Vora Graph Engine  
Vora Document Store  
Distributed Log  
Docker Registry



Requirements  
CPU  
Memory  
Disk

Preconfigured  
Building Blocks

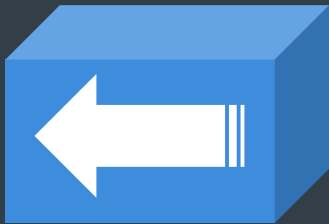


Lenovo Sizing Experts

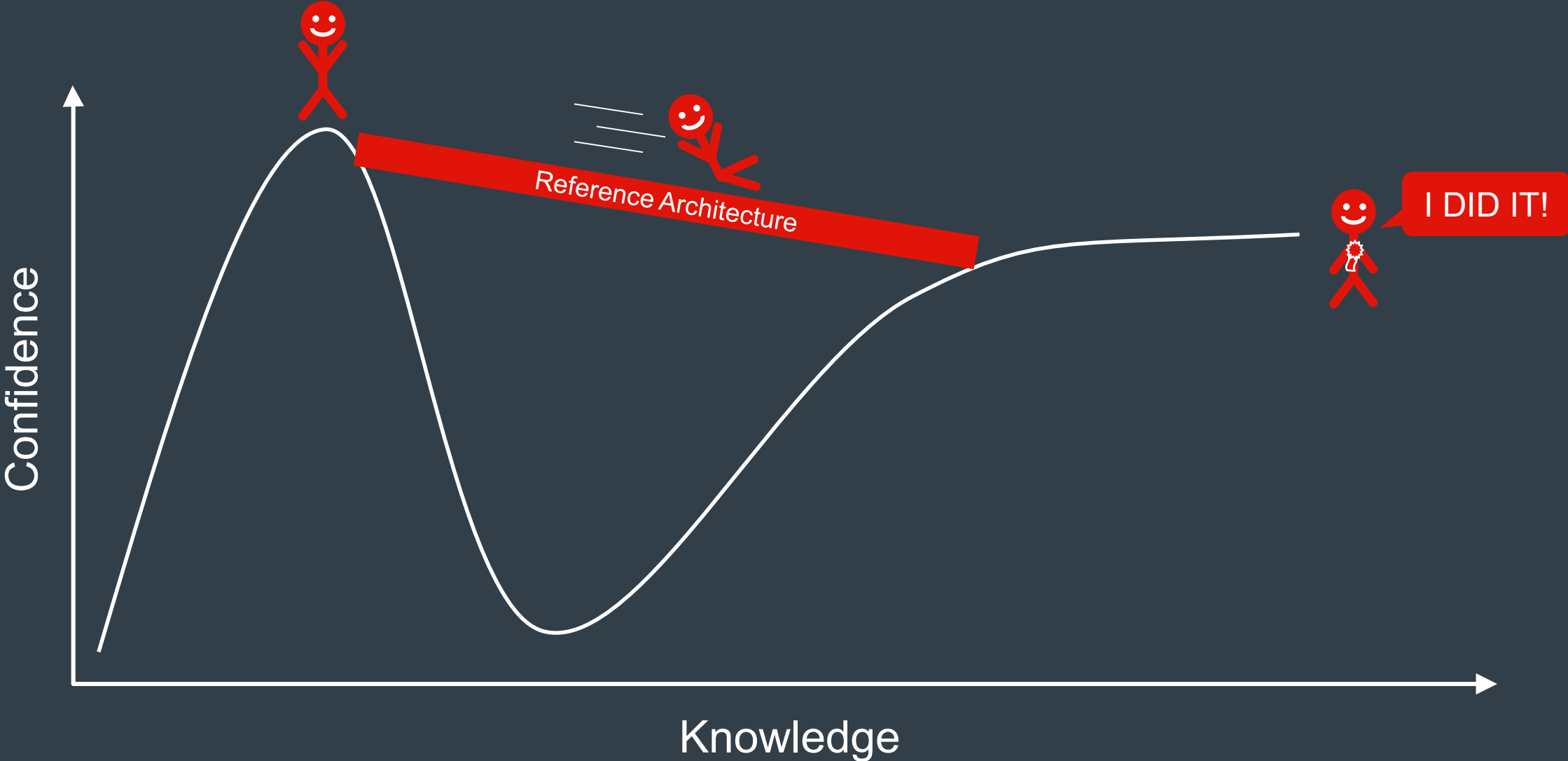


Infrastructure  
Configuration

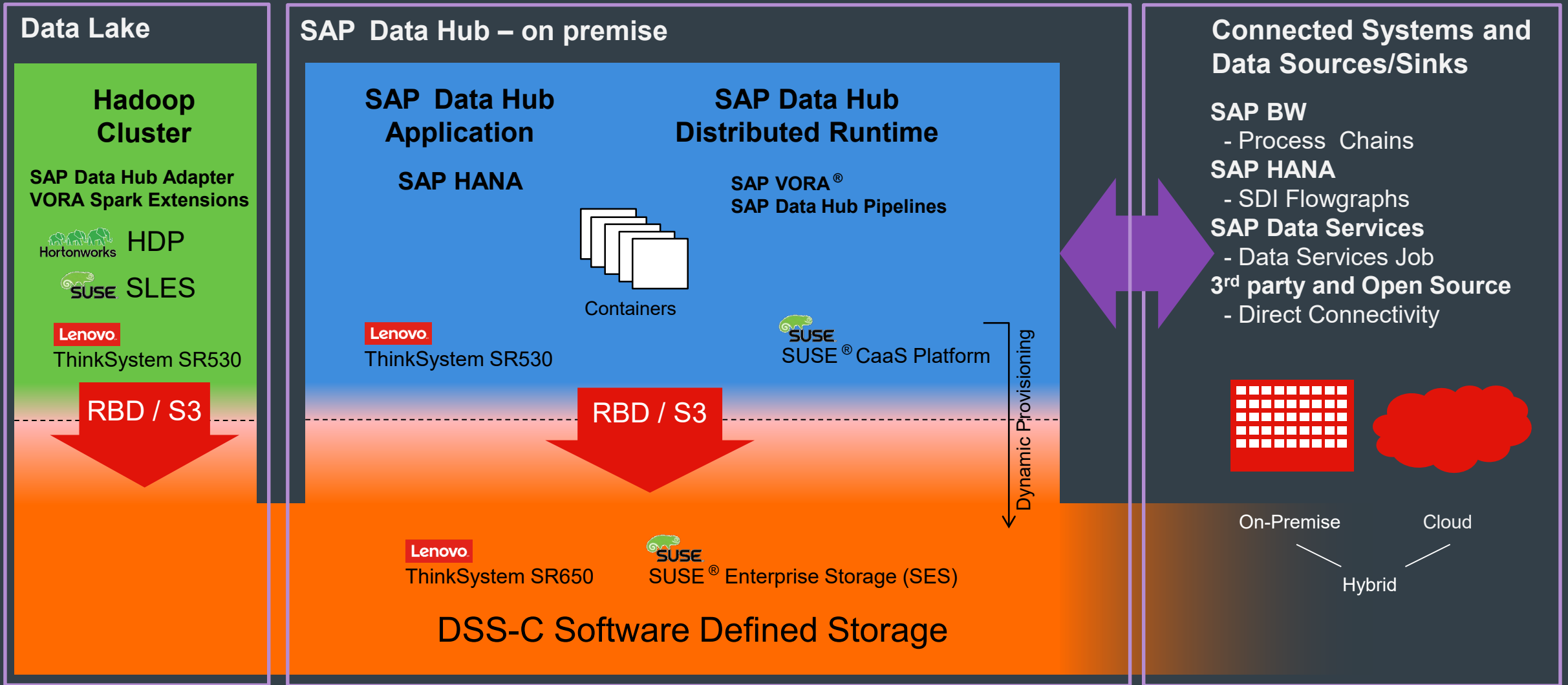
Data Hub  
Units  
(SAP Licensing)



# Why a Reference Architecture?



# Lenovo Solution for SAP Data Hub



\* All information provided here is subject to change and may be changed by Lenovo at any time, for any reason, without notice. It is not a commitment, promise or legal obligation to deliver.

# Lenovo Solution for SAP Data Hub

Simplify Deployment, Accelerate Implementation

- ✓ Engineered solution, fit to purpose
- ✓ Flexible building block approach
- ✓ Scalable Architecture
- ✓ SAP<sup>®</sup>, SUSE<sup>®</sup> and Lenovo<sup>™</sup> Best Practices built in
- ✓ Easy deployment





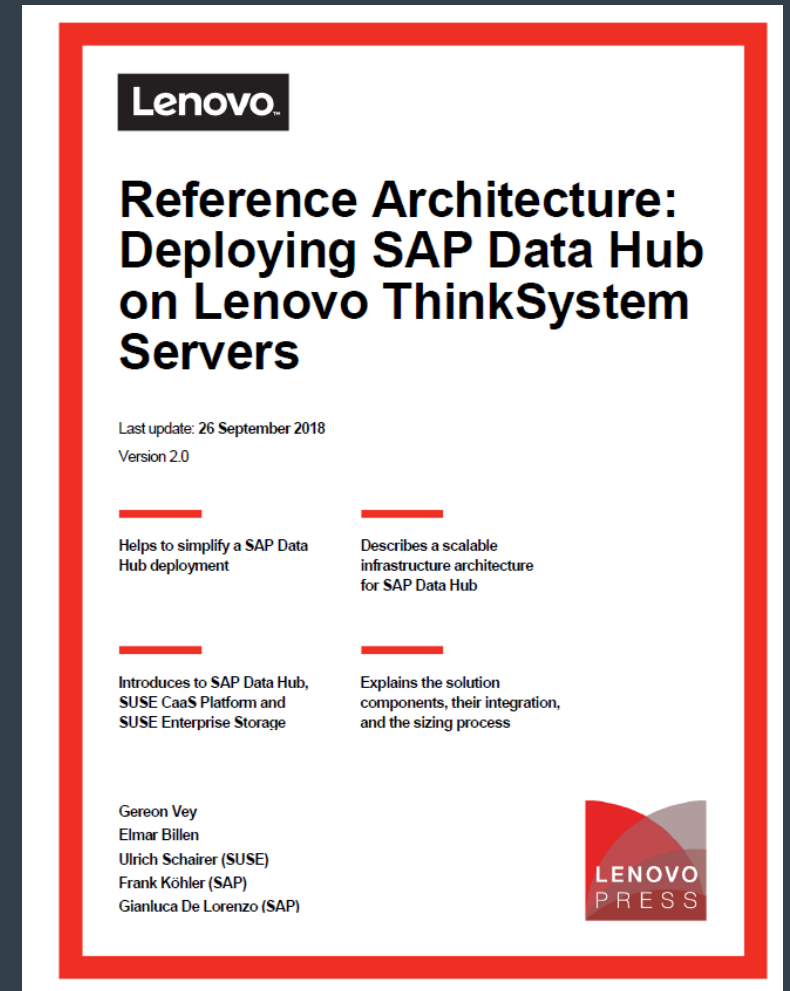
# Reference Architecture

Find out more:

Reference Architecture:  
Deploying SAP Data Hub  
on Lenovo ThinkSystem Servers

Download from Lenovo Press

<https://lenovopress.com/lp0897>



Q & A

# thanks.

Different is better

Lenovo

Powered by Intel® Xeon®  
processor Scalable family

