Practical Architecture for SAP Data Hub on Premise

Nathan Saunders, SAP North America Alliances
Jose Betancourt, SUSE Global Alliances Director
Data is kept in Silos across the Enterprise

- Plain Files
- IoT
- Mobile
- Cloud Storage
- Databases
- Enterprise Applications
- Data Warehouses
- Hadoop Data Lakes
Unlock the Value of your Data

Plain Files

Enterprise Applications

Databases

IoT

Mobile

Cloud Storage

Data Warehouses

Hadoop Data Lakes

Plain Files

Databases

Enterprise Applications

IoT

Mobile

Cloud Storage

Data Warehouses

Hadoop Data Lakes
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

All Devices

SAP® Solutions (such as SAP S/4HANA®), ISV and custom applications

SAP HANA®
- Application services
- Database services
- Integration services

SAP Data Hub
- Connectivity, process definition, orchestration

SAP Vora®
- Data-processing engines

Distributed Clusters
- Hadoop, S3, and more
What does SAP Data Hub actually do?

- **Pipeline**: Connect the Enterprise data silos in Real time or batch.
- **Orchestration**: Using workflow to manage event driven actions.
- **Ingestion**: EIM tool capability for discovery, transformation and data quality tools.
- **Control**: Strong tools around security, auditing and governance.
- **Data Discovery**: Create metadata profiles for data held in your silos for easier use.

End-to-End Data Management
Why a Reference Architecture?

- EASY!
- Mt Stupid
- Valley of Despair
- Slope of Enlightenment
- Plateau of Sustainability
- This will never work...
- I DID IT!
What do we want?

Scalability

Flexibility

Ease of Deployment

Performance

Fault Tolerance

Security

Compact Footprint

2019 Lenovo. All rights
SAP Data Hub Architecture

**SAP Data Hub Foundation**

- **Tenant Applications / Services** (Managed by System Management)
  - Pipeline Modeler
  - Flow Agent
  - Metadata Explorer
  - Spark on Kubernetes
  - Connection Management

- **Diagnostics**
  - VORA Database
  - System Management

- **SAP HANA** (internal)

**Kubernetes Cluster**

**Data Stores**
- Cloud / On-Premise
  - Hadoop Cluster
    - VORA Spark Extensions
    - HDFS / Spark
  - Cloud Stores
    - AWS S3, GCP GCS, Azure ADL & WASB

**Docker Registry** (secure)
SAP Data Hub implementation components

- SAP Data Hub Application
- SAP Data Hub Distributed Runtime
- Kubernetes
- OS
- Server hardware
- Fabric for Networking & Storage
- Storage hardware
- Hadoop and other components
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

2019 Lenovo. All rights
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

2019 Lenovo. All rights
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

2019 Lenovo. All rights
SUSE Linux Enterprise for SAP Applications

- 24x7 Priority Support for SAP
- Extended Service Pack Support
  - 18 Month Grace Period
- Page Cache Management
- Installation Wizard
- SAP HANA Firewall
- SAP HANA Resource Agents
- SUSE Linux Enterprise Server
- SLE High Availability
  - SAP HANA & SAP NetWeaver
- Extended Service Pack Support
  - 18 Month Grace Period

2019 Lenovo. All rights
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

- **SAP Data Hub Application**
- **SAP Data Hub Distributed Runtime**
- **SUSE Container as a Service Platform**
- **Hortonworks Data Platform** (optional)
- **SUSE Linux Enterprise Server**
- **Lenovo ThinkSystem SR530**
- **25GbE infrastructure**
  - **Lenovo Rack Switch NE2572**
- **Software Defined Storage**
  - **Lenovo ThinkSystem SR650 with SUSE Enterprise Storage**
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

2019 Lenovo. All rights
**Physical View**

**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

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

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

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

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

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

Powered by Intel® Xeon® processor Scalable family

2019 Lenovo. All rights
# 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
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

Lenovo Sizing Experts

Preconfigured Building Blocks

Data Hub Units (SAP Licensing)

CPU

Memory Disk

Lenovo Sizing Experts

Infrastructure Configuration

2019 Lenovo. All rights
Why a Reference Architecture?

Confidence

Knowledge

Reference Architecture

I DID IT!
Lenovo Solution for SAP Data Hub

**Data Lake**
- Hadoop Cluster
  - SAP Data Hub Adapter
  - VORA Spark Extensions
- SAP HANA
- SUSE SLES
  - Lenovo ThinkSystem SR530
  - RBD / S3

**SAP Data Hub – on premise**
- SAP Data Hub Application
  - SAP HANA
  - SAP VORA®
  - SAP Data Hub Pipelines
  - Containers
- Lenovo ThinkSystem SR530

**SAP Data Hub Distributed Runtime**
- SUSE® CaaS Platform
- Dynamic Provisioning

**Connected Systems and Data Sources/Sinks**
- SAP BW
  - Process Chains
- SAP HANA
  - SDI Flowgraphs
- SAP Data Services
  - Data Services Job
  - 3rd party and Open Source
  - Direct Connectivity

**DSS-C Software Defined Storage**
- On-Premise
- Cloud
- Hybrid

---

*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®, SUSE® and Lenovo™ Best Practices built in
- Easy deployment

Reduced Risk, Faster time to value

I DID IT!
Reference Architecture

Find out more:

Reference Architecture: Deploying SAP Data Hub on Lenovo ThinkSystem Servers

Download from Lenovo Press
https://lenovopress.com/lp0897
thanks.

Different is better