Achieve Low TCO and Enterprise Resiliency for SAP Data Hub and SAP HANA®

Gereon Vey – Lenovo
José Betancourt – SUSE
SAP Data Hub in a nutshell is the “umbrella” that simplifies the integration of all the solutions, data, and processing involved in SAP’s overall approach to Big Data. HANA is the well-known, in memory database for SAP applications. SAP Vora enables the management of Big Data on clusters and SAP Data Hub, which streamlines the use of SAP Vora, connects to additional sources and enables the definition and orchestration of processes in the distributed Big Data landscape.
While SAP Data Hub simplifies the integration of data across the enterprise, the implementation of SAP Data Hub is quite complex. There is a whole stack of supporting components needed to run SAP Data Hub:
The SAP Data Hub application runs on SAP HANA, the SAP Data Hub Distributed Runtime is containerized and needs a Kubernetes cluster to run on.
For an actual deployment, there are many decisions to be made regarding the underlying infrastructure
There is a choice of SAP HANA Platform edition or Express edition
There is at least a dozen of “leading” Kubernetes distributions, choice of operating systems, and the hardware has to be chosen and sized to the expected workload.
While having choices is generally a good thing, here it leads to
the risk of not getting it right the first time and thus derailing your SAP Data Hub implementation project. In order to address this, Lenovo has created a Reference Architecture for SAP Data Hub, jointly with SAP and SUSE.
There are three SUSE components providing the software-defined infrastructure for SAP Data Hub.

On the SAP HANA side, we have SUSE Linux Enterprise for SAP applications.

- The Linux development platform used by SAP.
- With product differentiators such as (HA SW and resource agents – documentation on HA scenarios, page cache management, etc.).
- Priority support with a single point of contact for escalations – SAP.
The second component is SUSE’s Container as a Service Platform.
SUSE CaaS Platform is an application development and hosting platform for container-based applications and services that enables you to provision, manage, and scale container-based applications and services, so you can focus on development of container applications and meet your business goals faster while saving costs in developing and maintaining container infrastructure.
The easy-to-use container platform includes all required components—orchestration with Kubernetes, microservices and container OS, and configuration—as one solution. With this unified platform, you can easily build microservices-based container applications that run on an agile container host OS optimized for containerized apps.
SUSE® Enterprise Storage, powered by Ceph, a highly scalable and resilient software based storage solution, enables organizations to build cost-efficient and highly scalable storage using commodity off-the-shelf servers and disk drives. It is self-managing and delivers storage functionality comparable to mid- and high-end storage products at a fraction of the cost.

Key points to highlight:
- Ability to provide file, block, and object with a single storage solution.
- Adaptable – not only via gateways, but leveraging different HW configuration scenarios.
- Partnership with Lenovo already yielded reference configurations for SAP HANA (in use at SAP and Lenovo as actual customers)
Leveraging SLES, SUSE CaaSP, and SES, Lenovo built a reference architecture for SAP Data Hub.

The compute layer [above the dashed line] consists of:
A Lenovo ThinkSystem SR530 running SLES and SAP HANA Express Edition for the SAP Data Hub Application
A cluster of SR530s running SUSE CaaSP for the SAP Data Hub Distributed Runtime

The storage layer [below the dashed line] consists of a cluster of Lenovo ThinkSystem SR650 servers forming a software-defined storage system using SUSE Enterprise Storage. It is used as a storage for SAP HANA, and to provide persistent volumes for containers running on SUSE CaaSP. These persistent volumes can be provisioned dynamically.
Lenovo Solution for SAP Data Hub – Outlook*

- **Data Lake**
  - Hadoop Cluster
  - SAP Data Hub Application
  - SAP HANA
  - SAP Data Hub Distributed Runtime
  - SAP VORA™
  - SAP Data Hub Pipelines
  - SAP Data Hub Adapter
  - VORA Spark Extensions

- **SAP Data Hub – on premise**
  - RBD/CephFS
  - Lenovo ThinkSystem SR530
  - SUSE® SLES

- **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**

* 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
thanks.

Different is better