Lenovo Solution for SAP Data Hub

A scalable landscape that meets data growth challenges

Manage data-driven processes for end-to-end Big Data scenarios

The Data Challenge

New data sources from new use cases such as IoT, video imaging, mobile devices and real-time monitoring are driving the exponential growth of data. Many enterprises are challenged to store the data as well as gain business insights and competitive advantage from the data. At the same time, corporate data landscapes are growing increasingly complex making it hard and costly to capture the maximum value from the available data.

Usually data is kept in silos across the enterprise, e.g. in enterprise applications, databases, plain files, Hadoop data lakes, data warehouses, or various forms of cloud storage. Combining data across those silos is needed to unlock its value, but this process is complex, time-consuming and therefore costly. Many of today's data integration tools are point-to-point, complex to use, and highly manual making it challenging to rapidly connect and implement desired data outcomes. The increased complexity of enterprise landscapes also makes it challenging to enabling appropriate and effective governance. In order to trust and rely on data accuracy, an end-to-end governance across all data sources is required otherwise acting on the data, either through analytical or operational applications using the data, is at risk.

Enterprise readiness of Big Data technologies is not a given. When trying to solve the complexity of the data landscape by simply storing all data in a Hadoop data lake, businesses often encounter limited governance, little automation of scheduling data processing, lack of common security and access management, and limited monitoring and tracing capabilities. Also, implementing Big Data initiatives and creating value from them requires very specialized skillsets. These specialized resources can be difficult to find and retain. Managing and processing data across silos is also a challenge, but it is also an opportunity to unlock the value of the data by combining data from different sources to create new insights.

Highlights
- Improves decision making whether data is in the cloud, on premise, in data lakes, or business systems
- Scales your data center landscape to enable new use cases and meet demanding data growth
- Prepares you for future information management needs with an extensible and open system

Built on server infrastructure rated #1 in reliability and customer satisfaction[1]

Figure 1: SAP Data Hub
Experience a simpler, more scalable approach to data operations and landscape management
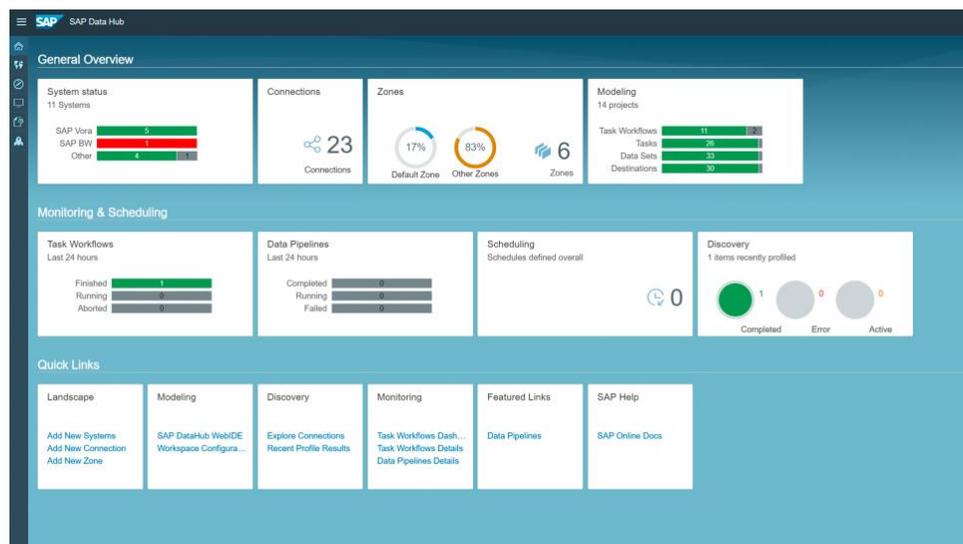
MANAGE DATA-DRIVEN PROCESSES FOR END-TO-END BIG DATA SCENARIOS

SAP Data Hub provides a simple, scalable approach to manage, integrate, process and govern data. It is a DataOps management solution that enables agile data operations across the enterprise that supports data sharing, pipelining, and governance of all data in the connected landscape. SAP Data Hub is an open-data architecture that works across Hadoop, data lakes, cloud object storage, relational databases, enterprise applications, and more.

SAP Data Hub helps organizations to better understand data sources, interconnections, quality and impact by providing a broad and detailed view of the entire data landscape. SAP Data Hub provides a single data management pane for data from various data sources, like Hadoop, cloud storage, SAP HANA, business applications, and more. This enables enterprises to discover new business opportunities, resolve emerging data issues, and ensure data flowing to where it needs to be.

Most of the data that SAP Data Hub will process typically comes from non-SAP systems. SAP Data Hub is designed to leave the source data in place (even if the aggregated data goes into SAP HANA).

Data Hub uses "push down" processing techniques to enable the source systems to process the data in place as much as possible before pushing results to SAP Data Hub.

Leaving data where it is, is important. Data movement is time-consuming and expensive. It creates superfluous copies of the data, which can be a compliance nightmare.

Pipeline processing is done on SAP Vora, the data processing platform based on Apache Spark. SAP Vora has the option to store the processed data in SAP HANA is available.

DataOps management
Experience a simpler, more scalable approach to data operations and landscape management with enterprise-spanning data governance and easier data integration.

Accelerate data projects
Expand data projects by easily creating powerful data pipelines in a single, visual design environment.

Data-driven applications
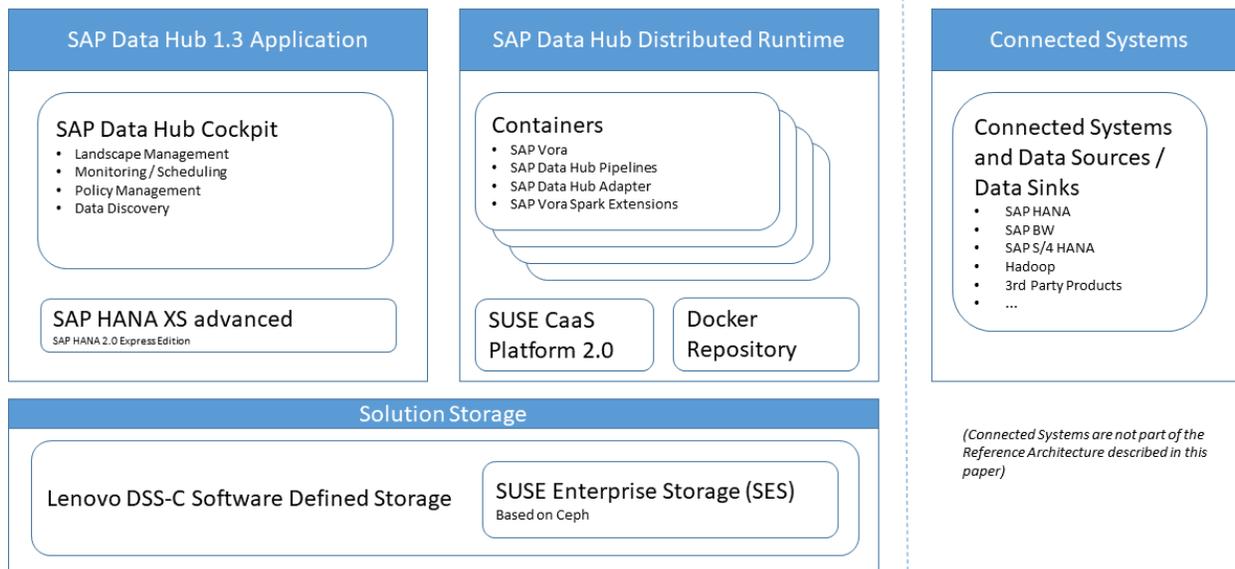Accelerate your business with fast pipeline results enabled by "push-down" distributed native processing.

Figure 2: Lenovo Solution for SAP Data Hub Architecture

SCALE YOUR DATA CENTER LANDSCAPE

Lenovo Solution for SAP Data Hub is a critical starting point to enable a successful Digital Transformation. The solution enables unified analytics of multi-layer, multi-vendor, multi-location environments (on-premises, cloud, hybrid cloud). In addition, it provides executive-level oversight of devices and data.

A complete SAP Data Hub implementation is composed of several components including the SAP Data Hub Cockpit, SAP HANA, SAP Data Hub distributed runtime connected systems and solution storage.

The Lenovo Solution for SAP Data Hub is an integrated end-to-end

solution built on selected, proven-to-work components. A flexible building block approach guarantees a high scalability, from proof of concept to large production implementations. SAP, SUSE and Lenovo Best Practices are built into the solution for best availability, performance results and and predictable outcomes.
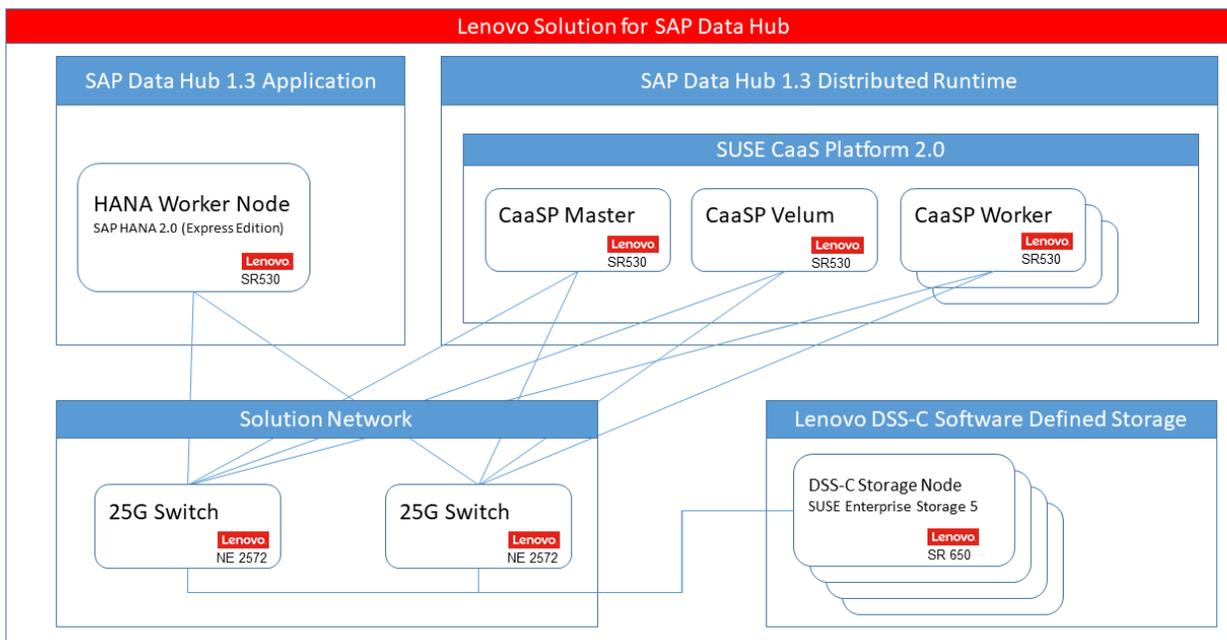

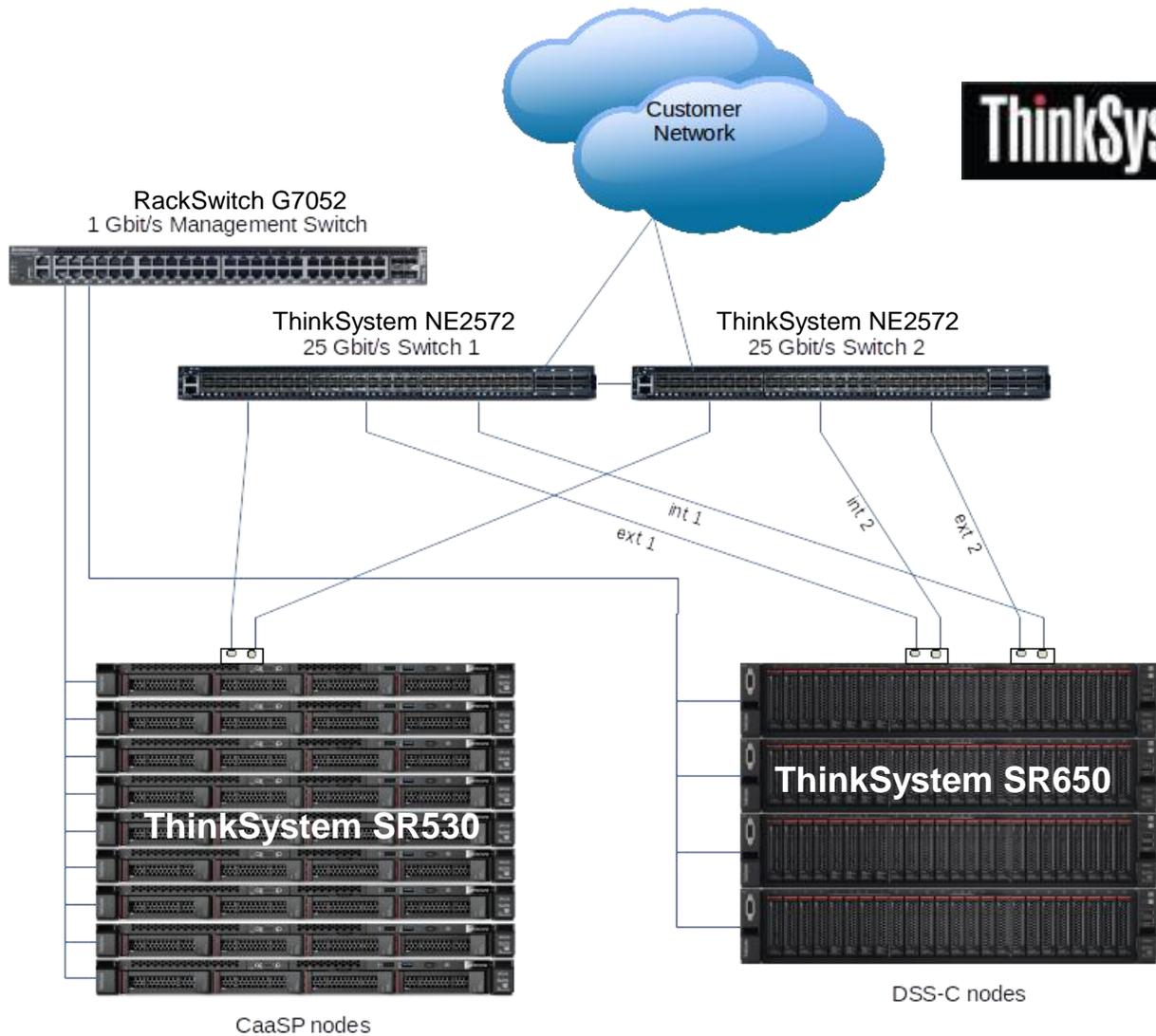**Figure 3: Lenovo Solution for SAP Data Hub Architecture**

Figure 4: **Lenovo Solution for SAP Data Hub Network Architecture** .

AVOID THE IMPACT AND RISK OF DOING NOTHING

Lenovo Solution for SAP Data Hub solution components include:

- **SAP Data Hub Application tier** is based on the SAP HANA Extended Application Services (XS) running on SAP HANA.
- **SAP Data Hub Distributed Runtime** is a fully distributed, containerized solution running on a Kubernetes infrastructure
- **SAP Vora** is a distributed database, which stores and processes structured data, graph data and semi-strucured document in-memory and on disk,

which provides enriched interactive analytics on data stored in Hadoop or other sources. It runs containerized in the Kubernetes infrastructure and provides a bi-directional interface between data from SAP Data Hub data sources to Hadoop.

- **SAP Data Hub Pipelines** are the connectors between the various SAP Data Hub data sources and provide reusable, configurable operations to process data from the various sources.

SAP Data Hub Pipelines run containerized in the Kubernetes infrastructure.

- **SAP Data Hub Adapter** serves as the central communication endpoint for operations performed from the Data Hub application on the Data Hub Distributed Runtime. It runs containerized in the Kubernetes infrastructure.
- **SAP Data Hub Vora Spark Extension** provides access to SAP Vora data sources from a Spark environment. It is installed on each Spark worker node in a Hadoop cluster.

A SUSE CaaS Platform cluster consists of several types of nodes:
- Administration Node
- Kubernetes Master Node
- Kubernetes Worker Nodes
- Private Docker Registry

SUSE Enterprise Storage (SES) is a commercially supported distribution of Ceph, an enterprise grade, scale-out storage solution. Ceph is a distributed object store which provides excellent performance, scalability and reliability.

The Lenovo Distributed Storage Solution for CEPH (DSS-C) provides a scalable platform for SES based on a scalable server building block. The storage capacity of the SES solution can be expanded easily by integrating additional storage client nodes to the cluster.

The Lenovo Solution for SAP Data Hub is deployed on Lenovo ThinkSystem platforms:

- **Lenovo ThinkSystem SR530** Server – SAP Data Hub application and distributed runtime
- **Lenovo ThinkSystem SR650** Server – DSS-C storage nodes
- **Lenovo ThinkSystem NE2572 Rackswitch** – Dedicated SAP Data Hub network
- **Lenovo RackSwitch G7052** – Management network

Lenovo Professional Services provide end to end lifecycle management as well as deployment, maintenance and managed services for SAP HANA and big data applications.

Ensure Business Transformation Success, Lenovo Solution for SAP Data Hub:

- Improves decision making whether data is in the cloud, on premise, in data lakes, or business systems
- Enables new business use cases
- Prepares you for future information management needs with an extensible and open system
- Scales to meet ever growing enterprise and big data requirements based a server infrastructure rated #1 in reliability and customer satisfaction[1]

#1 in **x86**
Reliability
In the latest ITIC reliability survey 2017 – 2018, **Lenovo** again was the **#1 x86 vendor** for five years running*

Future Data Expansion

CaaSP Nodes
SAP HANA
ThinkSystem
SR530
G7052 1 GBit/s Mgmt
NE2572 25 GBit/s
DSS-C Storage
ThinkSystem
SR650

1U 2U 3U 4U 5U 6U 7U 8U 9U 10U 11U 12U 13U 14U 15U 16U 17U 18U 19U 20U 21U 22U 23U 24U 25U 26U 27U 28U 29U 30U 31U 32U 33U 34U 35U 36U 37U 38U 39U 40U 41U 42U
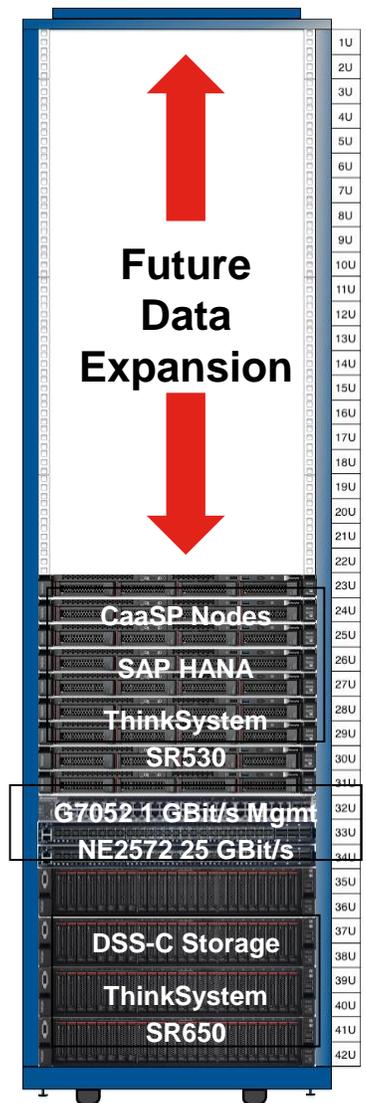
Figure 5: Lenovo Solution for SAP Data Hub Rack Implementation

Drive Successful Digital Transformation

Why Lenovo

Lenovo is a leading provider of x86 servers for the data center. Featuring rack, tower, blade, dense and converged systems, the Lenovo server portfolio provides excellent performance, reliability and security for SAP applications. Lenovo has shipped more than 10,000 SAP HANA systems and also offers a full range of networking, storage, software, solutions, and comprehensive services supporting business needs throughout the IT lifecycle. With options for planning, deployment, and support, Lenovo offers expertise and services needed to deliver better service-level agreements and generate greater end-user satisfaction.
.

For More Information

To learn more about the Lenovo Solution for SAP Data Hub contact your Lenovo sales representative or Business Partner and visit: www.lenovo.com/sap