ABSTRACT

Today’s businesses are generating far more data than ever before. Now there are also more ways to store and consume it. SAP Data Hub provides a simple, scalable approach to manage, integrate, process and govern data.

This document describes the reference architecture for SAP Data Hub (version 2) deployed on Red Hat OpenShift Platform, using Red Hat Ceph Storage Platform. The Dell EMC, Red Hat and SAP teams worked together on the architectural vision and joint engineering effort that gave rise to this white paper. This white paper is intended to offer some insight into the deployment of SAP Data Hub on-premises, using Dell and Red Hat products.

The intended audience of this document is IT professionals, technical architects, sales engineers and consultants to assist in planning, designing and implementing SAP Data Hub.

May 2019
# TABLE OF CONTENTS

**SAP DATA HUB** .......................................................... 1
  Introduction ....................................................... 1
    Functionality .................................................. 2
  Use cases ........................................................ 2
    Internet of Things ........................................... 2
    Data science and machine learning data management ........... 3
    Intelligent data warehousing .................................. 3
    Governance with data cataloging ............................ 3

**KEY SOLUTION COMPONENTS** .................................. 3
  SAP HANA .......................................................... 4
  SAP Data Hub .................................................... 4
  Red Hat ........................................................... 4
    Red Hat SAP Data Hub architecture ........................ 4

**SOFTWARE SUMMARY** ........................................... 5

**SYSTEM ARCHITECTURE OVERVIEW** ......................... 7
  Network architecture ......................................... 8
  Load balancing ................................................ 8

**HARDWARE DETAILS** ............................................. 9
  For production-ready configuration .......................... 9
  Hardware configuration ....................................... 10
  Networking overview .......................................... 11
  Management network ......................................... 12

**KEY TAKEAWAYS** ................................................ 12

**LEARN MORE** .................................................... 12

---

The information in this publication is provided “as is.” Dell Inc. makes no representations or warranties of any kind with respect to the information in this publication, and specifically disclaims implied warranties of merchantability or fitness for a particular purpose.

Use, copying and distribution of any software described in this publication require an applicable software license.

Copyright © 2019 Dell Inc. or its subsidiaries. All Rights Reserved. Dell, Dell Technologies, EMC and other trademarks are trademarks of Dell Inc. or its subsidiaries. Intel, the Intel logo and Xeon are trademarks of Intel Corporation in the U.S. and/or other countries. Other trademarks may be the property of their respective owners.

Dell Technologies believes the information in this document is accurate as of its publication date. The information is subject to change without notice.

Published in the USA 05/19.
INTRODUCTION

Today, corporate data landscapes are growing increasingly complex. The volume of data is growing exponentially, driven by the massive expansion of unstructured data, such as that from the Internet of Things (IoT), images and social data. Moreover, data is stored in multiple locations — on-premises, in the cloud and on edge devices. Existing systems and processing methods, which were built primarily for managing structured transactional data, are typically point to point and highly manual. New types of processing techniques, such as machine learning, require specialized, expensive resources to execute at scale. As a result, companies are challenged to deliver actionable information instantly and automatically where it’s needed.

The SAP Data Hub solution provides data pipelining, governance and landscape management of diverse data across distributed data landscapes. It refines and distills a variety of data using automated data pipeline operations for scale and efficiency. A metadata catalog helps provide the governance to ensure the right data gets to the right user, in the right context, at the right time — for data-driven empowerment. It enables access to distributed landscapes of diverse data, in a hybrid cloud model, giving you centralized visibility and control of your data.

SAP Data Hub provides a single solution to quickly create powerful, scalable data pipelines that connect data sources without moving the data, to enrich the data with context and to ensure data quality, integrity and security. These capabilities dramatically decrease the time to deliver intelligent data your business can use to optimize processes, respond to new information and events, or expand your company’s portfolio of digital offerings. Modeling data pipelines can activate powerful libraries for computation or machine learning; for example, to rapidly connect data of a wide variety of types, such as IoT, social media, customer and product information. Through modeling, you can also leverage existing processing investments, such as capabilities in the SAP HANA® business data platform, Apache Hadoop or Apache Spark. Further, SAP Data Hub is taking advantage of container technologies, such as Docker, and container management tools, such as Kubernetes. You
can copy, modify and reuse data pipelines to accelerate the development and deployment of data-driven applications. By leveraging Red Hat® OpenShift® as the underlying enterprise-grade container application platform, organizations can further decrease the time to value for their SAP Data Hub deployment and immediately benefit from a best-of-breed Kubernetes platform and best practices jointly developed by SAP and Red Hat.

SAP Data Hub provides centralized orchestration, pipelining and governance capabilities — a unified view and data operations management of the complex data landscape. A unique feature of the solution is its ability to leave data where it is, which means it does not have to be moved to a single massive data store. This provides advantages in terms of ease of management and speed of data-pipeline execution, and you can take advantage of your existing data stores and processing capabilities. SAP Data Hub does not offer its own persistent data storage; rather, it is a platform to orchestrate and process data between existing data stores.

**FUNCTIONALITY**

**All-in-one data orchestration** enables you to accelerate efficient delivery of intelligent data with innovative data pipelines that enable smart customer experiences. These capabilities include:

- GUI-based data pipeline workflow modeling
- Efficient data preparation and quality assurance
- Effective raw data enrichment using business data
- Powerful distributed data pipeline execution

**Effective data governance** allows you to compose pervasive data consumption strategies, gain complete visibility into all your data, and securely consume and share relevant data with metadata policy management for superior people engagement. Through improved governance, you can:

- Manage and catalog metadata for policy management across disparate systems
- Enable self-service data preparation
- Profile, view and expose the data of all connected systems

**Trusted data landscape management** empowers you to harness the value of exponentially growing data and expand your product and services portfolio with flexible open data landscape management. This includes:

- Extensive data-ingestion capabilities for a variety of data types and sources
- Native and remote connectivity to diverse landscapes
- Detailed policy management and security
- User-friendly graphical dashboard for monitoring and deployment

**USE CASES**

**INTERNET OF THINGS**

Tackle the challenge of understanding vast quantities of raw data and events from disparate, semi-structured sources. SAP Data Hub solves the challenge of capturing, structuring the unstructured and then analyzing data from distributed heterogeneous environments spanning messaging systems, cloud storage systems and enterprise applications. It enables enterprises to ingest, integrate and process sensor data, correlating it to structured application data to put it in full business context, and supporting a variety of modern and advanced processing paradigms to get actionable insights. These insights can be used to drive and influence intelligent processes.
**DATA SCIENCE AND MACHINE LEARNING DATA MANAGEMENT**

SAP Data Hub provides capabilities to ingest data of a multitude of types (such as sensor data, imagery and other forms of structured and unstructured data) and use machine learning to analyze patterns. SAP Data Hub provides one unified tool to process machine learning models and advanced analytics algorithms on any mix of engines, whether from SAP (predictive analysis library on SAP HANA, SAP Leonardo Machine Learning capabilities and so on) or from third-party providers (such as Python, R, Spark and TensorFlow). It helps data scientists and data architects to streamline data science projects, with one single, overarching tool delivering ingestion, preparation and processing across all kinds of disparate data, different engines and disparate processing paradigms. This finally allows enterprises to industrialize advanced analytics and machine learning in a cost-effective, time-effective and reliable way. Use cases such as smart energy management, transaction fraud detection and customer churn can now employ all trusted data for better machine learning.

**INTELLIGENT DATA WAREHOUSING**

Expand beyond traditional, monolithic data warehousing by orchestrating data, on premises or in the cloud, across data warehouses, data lakes, data marts and enterprise applications. Combine structured and unstructured data and process data where it resides within data pipeline applications. SAP Data Hub enables your organization to quickly acquire new data sources with previously siloed data, combining all types of sources in a governed way and streamlining a large variety of processing on them. For scenarios such as customer behavior analysis, SAP Data Hub brings the power of intelligent data.

**GOVERNANCE WITH DATA CATALOGING**

A variety of disparate data assets, structured and unstructured, distributed across different cloud and on-premises locations, are becoming more and more crucial to enterprises aiming to become data-driven businesses. But governance capabilities struggle to keep pace, and are often still siloed, covering only partial views. This creates a challenge in sharing the intelligence with all the consumers of information. With SAP Data Hub, you can enable consistent and proper handling of data quality, integrity and security across the enterprise. Profile, view and expose the data of all connected systems. And you can architect data models more quickly in a highly visual environment.

---

**Figure 2. SAP Data Hub architecture**
KEY SOLUTION COMPONENTS

SAP HANA
SAP HANA is a database designed to handle big data by processing transactions and analytics in-memory on a single data copy. It combines powerful data transformation with analytical functionality to deliver real-time insights from live data or analytical insights beyond the obvious.

SAP DATA HUB
One component of SAP Data Hub is SAP Vora, which provides interactive analytics on stored data from various sources, including traditional storage, Hadoop and Amazon Simple Storage Service (Amazon S3). Its specialized engines provide high-performance distributed analysis tailored by data category — such as relational data, graph data, semistructured JavaScript Object Notation (JSON) documents and time series data.

With SAP Data Hub, you can easily package analysis from disparate data types with a common SQL interface, easy-to-use modeling tools and extensions of Apache Spark to include SAP HANA\(^\text{\textregistered}\) platform access. As a result, you can gain insight based on relevant information from across your data environment.

RED HAT

RED HAT SAP DATA HUB ARCHITECTURE

SAP Data Hub on Red Hat OpenShift Container Platform is a continuation of the strong collaboration between Red Hat and SAP to support a variety of enterprise applications and big data scenarios. Running on Kubernetes clusters and Linux\(^\text{\textregistered}\) containers powered by OpenShift, SAP Data Hub provides containerized analytics engines and services and collects big data via Apache Spark, Apache Hadoop and directly from cloud environments to use for actionable business intelligence. Red Hat OpenShift delivers a scalable platform...
upon which these capabilities can run, integrating containerized services and SAP Data Hub resources for a unified, flexible offering. Key features of the integrated offering include:

• On-demand in-memory big data analytics
• Easier management of big data analytics at scale
• Easier integration of SAP Data Hub with SAP HANA
• Better support for agile development around big data use cases

SAP Data Hub provides interactive analytics on stored data from various sources, including traditional storage, Hadoop, public cloud storage, or hybrid-cloud- or on-premises-based solutions. It helps enterprises easily package analysis from disparate data types with a common interface, easy-to-use modeling tools and access to SAP HANA. This can provide deeper insights based on relevant information from across a business’s big data landscape.

Red Hat OpenShift Container Platform unites developers and IT operations on a single platform to build, deploy and manage applications consistently across hybrid- and multi-cloud infrastructures. Red Hat OpenShift helps businesses achieve greater value by delivering modern and traditional applications with shorter development cycles and lower operating costs. Red Hat OpenShift is built on open source innovation and industry standards, including Red Hat Enterprise Linux and Kubernetes, and is trusted by companies of all sizes around the world.

SOFTWARE SUMMARY

This reference architecture includes the following software:

• Red Hat Enterprise Linux as the operating system for all nodes
• Red Hat OpenShift Container Platform as the container application platform
• Red Hat Ceph® Storage to store application data and container images
• Red Hat Ansible® Automation for provisioning resources, deploying applications, and configuring and managing infrastructure
• Integrated Dell Remote Access Controller 9 (iDRAC9) Enterprise for remote server administration
• Ansible playbooks and modules for Dell Networking OS10 and iDRAC9 Enterprise

Additionally, the Red Hat OpenShift Container Platform architecture uses the following software:

• Docker runtime to build, ship and run containerized applications
• Kubernetes to orchestrate and manage containerized applications
• Etcd as the key-value store for the Red Hat OpenShift cluster
• Open vSwitch (OVS) for software-defined networking (SDN)-specific functions in the environment
In addition, hardware load balancing should be included where necessary.

| **Red Hat OpenShift Container Platform** | Red Hat OpenShift Container Platform is an enterprise-grade application environment that supports all aspects of the development process in one consistent platform, across multiple infrastructures. It integrates the architecture, processes, platforms and services needed to build applications. Using the Kubernetes container orchestrator, Red Hat OpenShift Container Platform manages applications across clusters of systems running the Docker container runtime. Find more information about Red Hat OpenShift Container Platform at [www.redhat.com/openshift](http://www.redhat.com/openshift). |
| **Red Hat Ansible Automation** | Red Hat Ansible Automation is simple, agentless IT automation technology that can be used to provision resources, deploy applications, and configure and manage infrastructure. It provides a visual dashboard that lets you manage all aspects of your automated tasks, including role-based access control (RBAC), job scheduling and real-time job status updates. The automation in this reference architecture is based largely on Ansible playbooks due to their simplicity and extensibility. Find more information about Red Hat Ansible Automation at [www.redhat.com/en/technologies/management/ansible](http://www.redhat.com/en/technologies/management/ansible). |
This reference architecture uses three types of nodes: master, worker and storage.

**Master Nodes**
- OpenShift Container Platform master API
- Ceph RADOS Gateway (S3)
- Internal VIP

**Worker Nodes**
- OpenShift Container Platform
  - Master + Infrastructure
  - OpenShift Container Platform Router

**Storage Nodes**
- Ceph Object Storage Device
- Ceph Monitor/Manager/RADOS GW

**External Network**
- VIPs
  - Master for OpenShift Container Platform
  - Fallback for:
    - OpenShift Container Platform master API
    - Ceph RADOS Gateway (S3)
    - Internal VIP
- LB1
- LB2
- Internal VIP

**Jump Nodes**
- Ansible Controller
- Red Hat Enterprise Linux

**Figure 4.** System architecture overview
**Master nodes**

Master nodes perform control functions for the entire cluster environment. They have application programming interface (API), controller management and scheduler capabilities and are responsible for creating, scheduling and managing all objects specific to Red Hat OpenShift.

The Red Hat OpenShift Container Platform web console runs as a pod on the master. The static assets required to run the web console are served by the pod. Managed from a web browser, the web console user interface allows users to visualize, browse and manage the contents of projects.

To achieve a low-latency link between etcd and the Red Hat OpenShift master nodes, you can install an etcd key-value store on the master nodes. Dell EMC, Red Hat and Intel recommend that you run both the Red Hat OpenShift master nodes and etcd in highly available environments. As described in this reference architecture, it’s recommended to run three Red Hat OpenShift master nodes in conjunction with an external active-passive load balancer and etcd clustering functions.

Master nodes act in this reference architecture as infrastructure nodes running a range of services, including an internal, integrated Docker registry, the L7 Load Balancer router and the management interface. Different components of OpenShift Container Platform react to new images, creating new builds and deployments. L7 Load Balancer provides routing functions for Red Hat OpenShift applications. It supports HTTP(S) traffic and Transport Layer Security (TLS)-enabled traffic via Server Name Indication (SNI).

**Worker nodes**

Worker nodes run containerized workloads. In this reference architecture, they run SAP Data Hub containerized applications like SAP HANA, SAP Vora and pipeline modeler. When a node fails, the affected pods are automatically rerun on the remaining nodes.

**Storage nodes**

Storage nodes provide persistent storage for the environment. You can manually or automatically create persistent volumes by storage classes. Two storage classes — one composed of SSD drives and another of NVMe drives — are specified in this reference architecture.

**NETWORK ARCHITECTURE**

- Red Hat OpenShift runs in a separate private VLAN.
- The only access is via a jump host (Red Hat Ansible Tower) and via load balancers.
- For security, the fewer servers with public/exposed IPs the better.

**LOAD BALANCING**

- Hardware routers/load balancers are recommended for production systems.
- Two load balancers, each with two interfaces:
- One on an exposed network — several VIPs (for master API, for router and RADOS Gateway)
- One on an internal network — one VIP is required there; the cluster nodes use this VIP as gateway

**HARDWARE DETAILS**

**FOR PRODUCTION-READY CONFIGURATION**

Figure 5 provides a rack-level view of the hardware used in this reference architecture. In addition to servers and switches, the rack also includes power distribution units (PDUs) and the necessary cables for network connectivity.

![Figure 5. A rack-level view of the hardware used in the architecture](image-url)
HARDWARE CONFIGURATION
Table 2 details the node configurations used in this reference architecture. Table 3 provides the full bill of materials, including rack infrastructure components, required to implement this reference architecture.

<table>
<thead>
<tr>
<th>Node</th>
<th>Quantity</th>
<th>Platform</th>
<th>Configuration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Master</td>
<td>3</td>
<td>Dell EMC PowerEdge R640 server</td>
<td><strong>Chassis and power</strong>&lt;br&gt;• 1 x 2.5-inch chassis with up to 8 hard drives and 3 PCIe slots&lt;br&gt;• Riser configuration 2, 3 x 16 low profile&lt;br&gt;• Dual hot-plug redundant power supply, 495 W&lt;br&gt;• 2 x C13 to C14 PDU style, 12 AMP, 6.5 ft (2 m)&lt;br&gt;<strong>Processor and memory</strong>&lt;br&gt;• 2 x Intel® Xeon® Gold 6138 processor, 2.0 GHz, 20C/40T, 27.5 MB L3 cache, 125 W&lt;br&gt;• 6, 32 GB RDIMMs (192 GB)&lt;br&gt;<strong>Storage</strong>&lt;br&gt;• Dell PowerEdge HBA330 Internal SAS HBA&lt;br&gt;• 2 x 240 GB Intel SSD SATA DC S4600 Series, software RAID 1 (container runtime ephemeral storage)&lt;br&gt;• Dell EMC BOSS controller card + 2 Dell/Intel 240 GB M.2 SATA SSD hard drive SSDSCKJB240G7R, configured in hardware as RAID 1&lt;br&gt;<strong>Connections and networking</strong>&lt;br&gt;• 1 x Intel X710 DP 10 GbE DA/SFP+, + i350 DP 1 GbE NDC&lt;br&gt;• 1 x Intel XXV710 DP 25 GbE SFP28 PCIe adapter, low profile&lt;br&gt;<strong>Software and management</strong>&lt;br&gt;• iDRAC 9 Enterprise</td>
</tr>
<tr>
<td>Worker</td>
<td>4</td>
<td>Dell EMC PowerEdge R640 server</td>
<td><strong>Chassis and power</strong>&lt;br&gt;• 1 x 2.5-inch chassis with up to 8 hard drives and 3 PCIe slots&lt;br&gt;• Riser configuration 2, 3 x 16 low profile&lt;br&gt;• Dual hot-plug redundant power supply, 495 W&lt;br&gt;• 2 x C13 to C14 PDU style, 12 AMP, 6.5 ft (2 m)&lt;br&gt;<strong>Processor and memory</strong>&lt;br&gt;• 2 x Intel Xeon Gold 6138 processor, 2.0 GHz, 20C/40T, 27.5 MB L3 cache, 125 W&lt;br&gt;• 6, 32 GB RDIMMs (192 GB)&lt;br&gt;<strong>Storage</strong>&lt;br&gt;• Dell PowerEdge HBA330 Internal SAS HBA&lt;br&gt;• 2 x 240 GB Intel SSD SATA DC S4600 Series, software RAID 1 (container runtime ephemeral storage)&lt;br&gt;• 2 x 960 GB Intel SSD SATA DC S4600 Series, software RAID 1 (OpenShift registry, logging and metrics)&lt;br&gt;• Dell EMC BOSS controller card + 2 Dell/Intel 240 GB M.2 SATA SSD hard drive SSDSCKJB240G7R, configured in hardware as RAID 1&lt;br&gt;<strong>Connections and networking</strong>&lt;br&gt;• 1 x Intel X710 DP 10 GbE DA/SFP+, + i350 DP 1 GbE NDC&lt;br&gt;• 1 x Intel XXV710 DP 25 GbE SFP28 PCIe adapter, low profile&lt;br&gt;<strong>Software and management</strong>&lt;br&gt;• iDRAC 9 Enterprise</td>
</tr>
<tr>
<td>Storage</td>
<td>6</td>
<td>Dell EMC PowerEdge R740xd server</td>
<td><strong>Chassis and power</strong>&lt;br&gt;• 1 x R740xd chassis with up to 24 x 2.5-inch hard drives including max of 12 NVMe drives, 2 x CPU configuration&lt;br&gt;• Riser configuration 6, 5 x 8 and 3 x 16 slots&lt;br&gt;• Dual hot-plug redundant power supply, 1100 W&lt;br&gt;• 2 x C13 to C14 PDU style, 12 AMP, 6.5 ft (2 m)&lt;br&gt;<strong>Processor and memory</strong>&lt;br&gt;• 2 x Intel® Xeon® Gold 6140 processor, 2.3 GHz, 18C/36T, 24.75 MB L3 cache, 140 W&lt;br&gt;• 12 x 16 GB RDIMMs 2,666 MT/s (192 GB)&lt;br&gt;<strong>Storage</strong>&lt;br&gt;• Dell PowerEdge HBA330 Internal SAS HBA&lt;br&gt;• 2 x 960 GB Intel SSD SATA DC S4600 Series, software RAID 1 (container runtime ephemeral storage)&lt;br&gt;• 22 x 960 GB Intel SSD SATA DC S4600 Series (persistent storage)&lt;br&gt;• Dell EMC BOSS controller card + 2 Dell/Intel 240 GB M.2 SATA SSD hard drive SSDSCKJB240G7R, configured in hardware as RAID 1&lt;br&gt;<strong>Connections and networking</strong>&lt;br&gt;• 1 x Intel X710 DP 10 GbE DA/SFP+, + i350 DP 1 GbE NDC&lt;br&gt;• 1 x Intel XXV710 DP 25 GbE SFP28 PCIe adapter, low profile&lt;br&gt;<strong>Software and management</strong>&lt;br&gt;• iDRAC 9 Enterprise</td>
</tr>
</tbody>
</table>
NETWORKING OVERVIEW

A 25 GbE network is the primary fabric for internode communication. Two Dell EMC S5148-ON switches provide data-layer communication, while one Dell EMC S3048-ON switch is used for out-of-band management. Figure 6 shows the network architecture.

This reference architecture uses virtual link trunking (VLT) to provide maximum availability. VLT is a layer-2 link aggregation protocol between end devices connected to two switches. It offers a redundant, load-balancing connection to the core network in a loop-free environment and eliminates the need to use a spanning tree protocol. VLT provides link connectivity between a server and the network via two different switches. VLT can also be used for uplinks between access or distribution switches and core switches.

Figure 6. Network architecture with virtual link trunking

Figure 7 shows the components of the Red Hat OpenShift Container Platform network environment and their logical architecture. Both physical 25 GbE interfaces are bundled together using 802.3ad link aggregation. The external network is presented as an 802.1Q virtual local area network (VLAN) tagged network from the upstream Dell EMC S5148-ON switch pair. The internal interface is presented as the untagged VLAN on the same switch pair to facilitate PXE booting. Link Aggregation Control Protocol (LACP) fallback is enabled so that nodes booting via Unified Extensible Firmware Interface (UEFI) can communicate with the provisioning system. This interface pair is denoted as Bond 0 in Figure 7.

Figure 7. Red Hat OpenShift Container Platform logical/physical network connectivity
All Red Hat OpenShift nodes are logically connected via the internal network on the same layer-2 broadcast domain. Additionally, OVS creates its own network for Red Hat OpenShift pod-to-pod communication. The Red Hat OpenShift OVS-multitenant plugin allows only pods with the same project namespace to communicate.

Use one of the spare 1 GbE LAN on Motherboard (LOM) interfaces on the bastion node to connect to the switch management ports from the bastion node. You can create an additional VLAN (i.e., VLAN 7) in the S3048-ON switch for this purpose. Connect all switch management ports and the 1 GbE LOM port on the bastion node to this VLAN and assign it to IP address 192.168.7.11 in the master node.

**MANAGEMENT NETWORK**

Use the 1 GbE management fabric (via the Dell EMC S3048-ON switch) for out-of-band server management and initial cluster deployment over the network from the master node. The iDRAC9 with Lifecycle Controller is embedded in all Dell EMC PowerEdge servers. It provides secure, remote server access for performing common management functions like configuration, operating system deployment, firmware updates, health monitoring and maintenance. It functions regardless of operating system state or hypervisor presence and is accessed via a dedicated 1 GbE port.

*Note: Do not connect the S5148F-ON management ports to this management network. Create a separate VLAN on the S3048-ON switch. Connect the S5148F-ON management ports and one of the 1-Gb LOM ports (i.e., em1) on the master node to this VLAN. This allows configuration and management of the S5148F-ON switches directly from the master node.*

**KEY TAKEAWAYS**

Today, corporate data landscapes are growing increasingly complex. The volume of data is growing exponentially, driven by the massive expansion of unstructured data, such as that from the Internet of Things, images and social data. To complicate matters, data is stored in multiple locations — on-premises, in the cloud and on edge devices. Existing systems and processing methods, which were built primarily for managing structured transactional data in enterprise systems, struggle with these new data challenges.

SAP Data Hub helps your organization take control of your data, wherever it is. It provides a simple, scalable approach to managing, integrating, processing and governing data. It’s a single solution designed to help organizations quickly create powerful, scalable data pipelines that connect data sources without moving the data, to enrich the data with context, and to ensure data quality, integrity and security. These capabilities can dramatically decrease the time required to deliver the intelligent data your business needs to optimize processes, respond to new information and events, and expand your company’s portfolio of digital offerings.

Together, Dell EMC, Red Hat and SAP can help you put your SAP Data Hub in place. We offer an end-to-end reference architecture backed by the leading-edge products, solutions and services you need to design, develop and deploy a robust on-premises SAP Data Hub tailored to the needs of your enterprise.

Ultimately, with an SAP Data Hub from Dell EMC, Red Hat and SAP, you will have the foundation you need to deliver actionable information instantly and automatically — wherever it’s needed.

**LEARN MORE**

To learn more:

- Contact your Dell EMC or Red Hat account representative.
- Visit [dellemc.com/sap](http://dellemc.com/sap).