FUJITSU Reference Architecture 2021 for SAP Data Intelligence 3.x

Contents

Introduction

Digital transformation requires platform solutions that support consistent and uniform management of entire SAP landscapes, including S/4HANA and SAP HANA, for every kind of IT provisioning model. In addition, organizations are facing more information than ever derived from multiple sources with the challenge to analyze and utilize the value of such data. Therefore, the main objective of this paper is to point out how business IT solutions developed by Fujitsu and its Technology Partners can provide a sustainable foundation for business continuity and future growth.

FUJITSU Limited

FUJITSU is the leading Japanese information and communication technology (ICT) company, offering a full range of technology products, solutions, and services. Approximately 132,000 FUJITSU people support customers in more than 100 countries. We use our experience and the power of ICT to shape the future of society with our customers.
www.fujitsu.com

FUJITSU is a world-leading IT services, solutions, and technology provider with extensive experience in designing, building, and deploying IT systems, services, and digital solutions for both public and private sector customers from retail and healthcare to financial services and automotive. For over 50 years, we have been at the forefront of technical innovation, investing $billions in research and development every year to ensure that we remain at the cutting-edge of business transformation. As experts in digital transformation, we are the partner of choice to keep you ahead of change in today's digitally enabled world. Find out more about our industry-specific IT services and solutions.
https://www.fujitsu.com/global/solutions/industry/

In today's highly competitive world, organizations across all sectors are turning to IT service providers to help them reduce costs, boost productivity and enhance performance. As a world-class end-to-end IT services provider, with over 80 years' experience working with organizations from all sectors around the globe, FUJITSU can provide the business IT solutions you need to improve efficiency, add value and lower costs. We design, develop, implement, manage and optimize access to systems and information to answer your business processing, application and infrastructure needs. Whether you are a private or public sector organization, or whether you want to run our solutions on your own hardware, or outsource your IT through us, FUJITSU has the expertise you need to overcome the business challenges you face. Explore our comprehensive portfolio of solutions which can be uniquely tailored to align with your individual business needs.
https://www.fujitsu.com/global/services/

FUJITSU Germany

As a wholly owned subsidiary of FUJITSU, FUJITSU Technology Solutions offers corporate customers in Germany a comprehensive portfolio of technology products, solutions and services that range from end devices to data center solutions, managed and maintenance services and cloud solutions to outsourcing. FUJITSU develops and manufactures notebooks, PCs, thin clients, servers, storage systems and mainboards and operates several highly secure data centers. With 8,000 channel partners in Germany, FUJITSU also has one of the most powerful partner networks in the industry.
https://www.fujitsu.com/de/

FUJITSU's Global SAP Partnership

In 1999, FUJITSU became the first global Technology Partner for SAP. Now, with a relationship lasting more than 4 decades, we have a wide offering of partner certifications that demonstrate the high-quality services we deliver to our customers.
We have expertise supplying end-to-end solutions that are supported by running SAP systems internally. In addition, we're also a strategic SAP customer, having implemented SAP services across our global organization. We have:

- almost 50 years of global partnership with SAP in technology and services
- more than 3,500 global SAP specialists
- more than 2,500 SAP systems and 330,000+ licenses in our responsibility
- more than 25 S4/HANA systems
- approximately 300 SAP service customers
- approximately 8,000 joint customer installations worldwide - supported by over 2,100 SAP application and solution experts
- european SAP License and Maintenance Intelligence
- business process integration and management for multi-cloud integration
- a leading position to deliver industrialized SAP services
- extensive experience in the use of S/4
- ability to deliver an end-to-end portfolio encompassing services, infrastructure, and re-selling of licenses

Fujitsu SAP solutions and support

At Fujitsu, we leverage the benefits of our global partnership with SAP for the benefit of our customers. We:

- provide an end-to-end SAP service and technology offering to our global customers
- provide the latest SAP solutions as well as first-hand support
- ensure our own product development remains closely in line with that of SAP
- enable organizations around the globe to utilize SAP business applications based on our cutting-edge, high-reliability platform products and consulting services
- with our combined offerings, we empower businesses to optimize their operations and boost their competitiveness

SAP Licenses

As a value-added reseller, we manage and optimize SAP licenses. We utilize individually customized hybrid licensing solutions with efficiency and On-Demand innovation. We provide, manage, maintain, and optimize SAP licenses based on:

- SAP Partner Managed Cloud subscription-based licensing for almost all On-premises SAP licenses
- SAP SaaS (Hybris, SuccessFactors, Ariba) reselling including value-added (e.g., Application Integration and Management)
- classical SAP license reselling
- blended IP reselling (e.g., SAP and FJ as non-SAP cloud solutions)

Leverage the benefits of our end-to-end SAP portfolio

As a trusted SAP Global Partner with over 40 years of experience, we have successfully helped thousands of customers around the globe to simplify and innovate operations as well as grow. Our expertise with SAP solutions, combined with our in-depth industry knowledge and technology experience, empowers you to achieve your business objectives and take the lead over your competition

We drive innovation in all areas around our three key pillars, namely Simplify, Transform and Digitalize, and provide a wide-ranging offering from cloud services via managed SAP operations to solutions and services for On-premises deployments. Our SAP support portfolio includes advanced technologies and service capabilities that deliver tailored future-proof solutions. In collaboration with our customers and specialist partners, we reduce complexity and optimize investment in SAP applications and the SAP HANA platform. The result is a more efficient enterprise, exceptional ROI, and better business outcomes.

By focusing on five principal areas, we help our customers digitally transform.

- simplify operations and the cloud for SAP solutions
- simplify business processes with SAP solutions
- transform to S/4 HANA
- transform to SaaS solutions to give a fuller picture
- digitalize and enhance the overall user experience across SAP offerings based on LEONARDO for SAP solutions

With our strategic FUJITSU SAP partnership, we harness and energize the power of innovation. We deliver a service offering that can provide everything you need to successfully shape your organization's digital transformation.

Our comprehensive range of services, solutions, and innovative platforms include SAP S/4HANA and SAP Leonardo. By working together to rationalize and reduce the complexity of your IT environment, we enable you to rapidly respond to changing market, business, and customer demands.

Our end-to-end SAP portfolio focuses on three key pillars, namely Simplify, Transform and Digitalize, which we use to provide a holistic view of your business needs and challenges. With our three-pillar approach, we identify the right solution that will enable your business to successfully transform in a digital reality.

For more information see https://www.fujitsu.com/global/services/application-services/enterprise-applications/sap/.

SAP Data Intelligence

IT experts are expecting that the annual data volume generated in 2025 could potentially sum up to 163 Zettabyte, which means 163 followed by 21 zeros (163 000 000 000 000 000 000 000 Byte). This is factor 10 of the data volumes generated in 2016, resulting in an annual growth rate of approximately 30%. Most of the data volume today is generated by private persons. Specialist are forecasting that about 60% of the data volume will be enterprise data in 2025, which is not distant future. Are you looking for a reliable solution that will help you to unlock one of the most valuable assets of your company, namely your business data? SAP Data Intelligence is a reliable platform supporting your company to conquer big data challenges.

SAP Data Intelligence is a data orchestration and management solution running on Kubernetes, which leverages open source and embeds machine learning capabilities. It was released by SAP in 2017 to deal with big data and complex data orchestration working across distributed landscapes and processing engines, enhancing developer productivity and extracting value from distributed data. SAP Data Hub was released in 2019 as SAP Data Intelligence, i.e., a managed service on SAP Cloud Platform. The name SAP Data Intelligence highlights the evolution of SAP Data Hub to operationalize data science and machine learning with the inclusion of the SAP Leonardo Machine Learning Foundation. SAP Data Intelligence provides all the integration, orchestration, metadata management, connectivity and rich services of SAP Data Hub with the services of SAP Leonardo Machine Learning in the cloud. SAP Data Intelligence 3.0, which extends the cloud version of SAP Data Intelligence to On-premises and BYOL (bring you own license model) deployments, is now generally available.

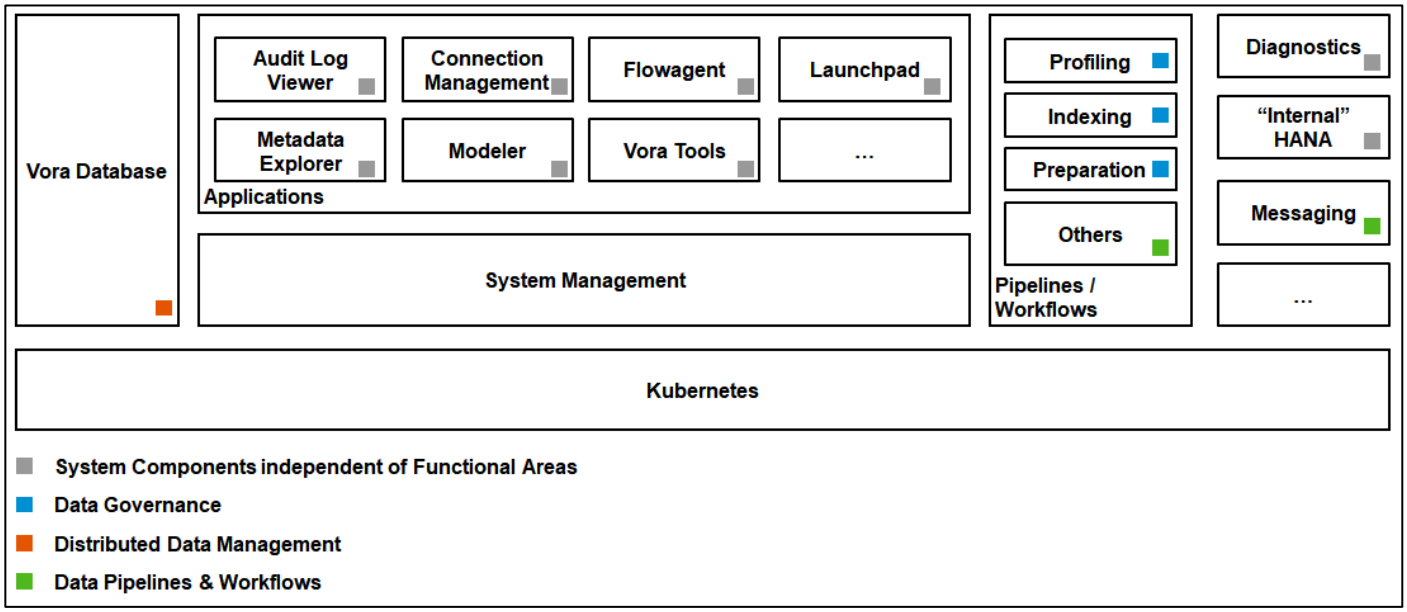


Figure 1: SAP Data Intelligence Architecture

In addition, it also comprises several basic services, such as SAP HANA, a messaging system (currently based on NATS and only used by data pipelines), a storage gateway (to abstract connectivity to storage), and a user account and authentication server (UAA). It is completely containerized and installed on Kubernetes. The capabilities of SAP Data Intelligence can be partitioned into the following functional areas:

- Data Governance: SAP Data Intelligence helps you to manage data across different systems by using the Metadata Explorer. It helps you to gather information about the location, attributes, quality, and sensitivity of data.
- Distributed Data Management: SAP Data Intelligence comprises a distributed database system for big data processing (SAP Vora Database). This allows you to store vast amounts of data and comes with support for various data types.
- Data Pipelines and Workflows: The Modeler allows you to build data pipelines and data workflows, which support a flow-based programming paradigm for data ingestion and transformation.

For more information refer to:
SAP Blogs | what-is-data-intelligence
SAP Blogs | Introduction to SAP Data Intelligence
SAP | SAP Road Map Explorer
SAP | SAP Data Intelligence - business cases
SAP Blogs | Data Intelligence – Considerations for Cloud or On-premises/BYOL installation
SAP Blogs | SAP Data Intelligence – connecting the Dots
SAP Blogs | Unified Data Integration for SAP
SAP Blogs | Debriefing: SAP Community Call "SAP Data Intelligence – Hybrid Data Management"
SAP Blogs | Want to Build a Resilient Business? Renew Your Focus on Connecting Data
SAP | SAP Data Intelligence - Machine Learning and Data Integration Services

SAP | Turn Data Chaos Into Data Value
SAP Data Intelligence | SAP Community
SAP Blogs | SAP Data Intelligence
SAP Blogs | SAP Data Intelligence Trial Edition 3.1
Docker | Docker Website
Kubernetes | Kubernetes Website

Fujitsu Reference Architecture for SAP Data Intelligence 3.x

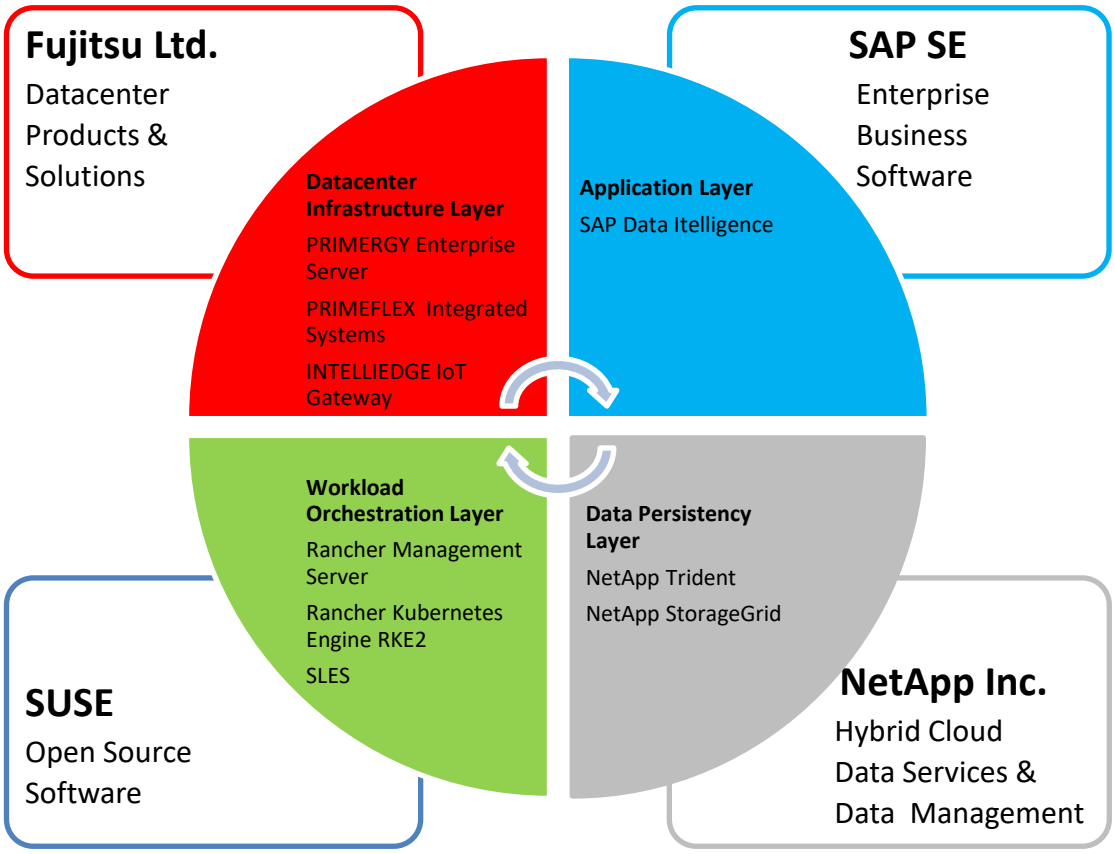


Figure 2: Fujitsu SAP Data Intelligence Reference Architecture Partners

In combination with SUSE Rancher Kubernetes Management Platform and NetApp Data Management Products, Fujitsu PRIMERGY servers and IT services form a reliable basis for the economical operation of your SAP Data Intelligence landscape.

The FUJITSU SAP Data Intelligence reference architecture enables proposals for customer-specific SAP Data Intelligence system landscapes to be created. The FUJITSU SAP Data Intelligence reference architecture is suitable for the operation of SAP Data Intelligence Platform. As part of our FUJITSU Integrated Systems PRIMEFLEX, FUJITSU also offers solutions, products and services for big data landscapes, which perfectly fit to the FUJITSU SAP Data Intelligence reference architecture. For detailed information please get in contact with Fujitsu | mailto Integrated Systems.

Four steps to your individual SAP Data Intelligence landscape

Step 1: Application Layer - SAP Data Intelligence Landscapes

This document only covers On-premises deployments. If you are looking for a deployment in the cloud, get in contact with your SAP sales representative. If you prefer an On-premises installation, continue by collecting relevant data to estimate the number of Kubernetes worker nodes.
How many Data Intelligence landscapes do you need?

- Development
- Quality Assurance
- Training
- Production & Staging (both clusters as identical as possible)
- other

Which function areas are you planning to use?

- Static components, User and Pipelines
- Data Governance
- Machine Learning
- Distributed Data Management

The most precise estimation would be to perform initial sizing as described in the SAP Data Intelligence sizing guide (SAP | SAP Data Intelligence 3.14 Sizing Guide). However, if you are at a very early stage of the project and some relevant data needed for the initial sizing are not yet known, you can use the SAP Data Intelligence Memory Calculator (SAP | On-premise SAP Data Intelligence Memory Calculator) to determine the number of worker nodes.

If you want to know how many services you can run on one worker node, perform an estimation using the numbers from the following table. When doing that take into consideration the following facts:

- if you are planning for non-productive use, the minimum number of worker nodes is 3
- 4 worker nodes are required for productive use
- the capacity of 1.5 worker nodes is reserved for SAP Data Intelligence demands and is not available for individual workloads
- the numbers are meant exclusively (not in parallel per single worker node)

Worker node size	32 GB/8 CPU	64 GB/16 CPU
Static components		
Tenant	6	12
Concurrent User	10	20
Pipelines		
Pipelines Small	16	32
Pipelines Medium	10	20
Pipelines Large	5	8
Data Governance		
Metadata Extraction Jobs	8	16
Data Profiling Jobs	8	16
Data Preparation Jobs	5	16
Machine Learning		
Jupyter User	1	3
Serving Model (2 GB RAM each)	16	32
Training Model (61 GB RAM)	0	1
Distributed Data Management		
Data Volume <= 5 TB	1	
Data Volume 5 – 30 TB	n/a	3 worker nodes required
Per Additional 10 TB	n/a	1 additional worker node

The following calculation scheme allows you to estimate the required number of worker nodes without any additional tools.
The minimum number of worker nodes is 4 which results in 256 GB of memory.

Specify the relevant type and estimated number of users. The number of concurrent users, not named users, is relevant. Typically, a ratio of 5:1 can be assumed for named users compared to concurrent users. Data Scientists working with Jupyter Notebooks should be counted both under total users as well as under Jupyter users.
Specify the type of activity (i.e., pipelines and jobs) you would like to run and enter the appropriate number. Remember to consider active concurrent pipelines/jobs at peak times (i.e., where you have the highest volume).
Examples for small pipelines:

- 1 source, <= 5 operators, only standard operators
- Data ingestion/extraction/load without transformations (<=2 systems involved)

Examples for medium pipelines:

- <=3 sources, <= 10 operators, standard operators and custom operators
- Data ingestion/extraction/load with structured transformations (<= 5 systems involved)
- Data transformation, anonymization, masking
- Data validation and rule execution
- Data stream processing
- System and process orchestration pipeline

Examples for large pipelines:

- >3 source, >10 operators, standard operators and custom operators, use of own libraries
- Data ingestion/extraction/load with complex transformations (> 5 systems involved)
- Machine learning model training or inference

Specify how many concurrent jobs you will have for serving and training of custom ML models. In addition, specify the type of compute node where you want to run these. Default configurations of 2 GB and 61 GB are given but these can be overridden according to individual needs and available hardware. For a rough estimate you can assume the following:

- Training on 61 GB nodes - 1 job for up to 5 models
- Serving on 2 GB nodes - 1 job for each model

	Capacity planning	Input units	GB	Sizing factor	Resource points (input units * sizing factor)
Workload Capacity	Static components				
	Static workload (fix value)	1	x	60	60
	Tenant	2	X	10	20
	Concurrent User	50	X	1	50
	Pipelines				
	Pipelines S	50	X	3	150
	Pipelines M	25	X	5	125
	Pipelines L	10	X	10	100
	Data Governance				
	Metadata Extraction Jobs	10	X	5	50
	Data Profiling Jobs	10	X	5	50
	Data Preparation Jobs	5	X	10	50
	Machine Learning				
	Number of Jupyter Users running 1 task/user	5	x	30	150
	Serving Model 2 GB RAM (# of Models * needed RAM in 2 GB RAM increments * Sizing factor)	5	2	3	15
	GB RAM per Training Model	1	61	95	95
	Distributed Data Management				
	Small Data volume < 5 TB	0		35	0
	Enhanced Data volume 5 TB – 29 TB	1		300	300
	Additional Data volume > 30 TB (per additional 10 TB = 1 * -200)	2		100	200
	Evaluation				
Evaluation	Sum of resource points				1415
	Number of 64 GB worker nodes (sum of resource points/100)				14

Note! This is only a rough estimation and should not be used instead of proper sizing.

Step 2: Orchestration Layer - Kubernetes Downstream User Cluster & Management Cluster

A brief overview about the SUSE Rancher products and usage scenarios are provided by:
SUSE Rancher | Rancher strengthens Kubernetes
SUSE Rancher | Rancher vs. RKE: What Is the Difference?

SUSE RKE itself comes in two versions - referred as RKE1 and RKE2. RKE2, as the most newest product that is also released in combination with SAP Data Intelligence, is the appropriate version to be used for your SAP Data Intelligence cluster considering the fact that operating system released versions of SUSE Linux Enterprise System (SLES) must be used. For more information refer to:
SUSE Rancher | RKE2 - How is this different from RKE or K3s?
SUSE | SAP Data Intelligence supported on Rancher Kubernetes Engine

SUSE Rancher RKE2 Downstream cluster

Additionally, to the worker nodes, Kubernetes requires nodes assigned for cluster management, named as Master nodes. Master nodes are running the control plane and the etcd services. The Control plane manages the worker nodes and the pods of the cluster. Etcd keeps the state of your cluster and is the most important component in your cluster. Kubernetes | Kubernetes Components
Depending on the system type of your SAP Data Intelligence cluster (DEV; QAS, PRD), different requirements and recommendations need to be fulfilled.
Each SUSE Rancher Kubernetes Engine Cluster (RKE2) running an SAP Data Intelligence System, referred to as a "Downstream Cluster", needs 3 types of nodes, namely:

- **Worker nodes**

Worker nodes run SAP Data Intelligence application workloads. The required number of worker nodes, RAM size and number of CPUs depend on your SAP Data Intelligence sizing and system requirements defined by SAP in:

Sizing and Installation Guides	SAP | SAP Data Intelligence Sizing & Installation Guides,
Product Availability Matrix (PAM)	SAP | SAP Product Availability Matrix,
and relevant SAP Notes	SAP | Note 2693555 - SAP Data Hub 2 and SAP Data Intelligence 3: Supported Remote Systems and Data Sources
	SAP | Note 2871970 - Prerequisites for installing SAP Data Intelligence 3
	SAP | Note 3038897 - Installation of SAP Data Intelligence 3 on Rancher Kubernetes Engine 2

Note! Use the most current versions and check for other relevant notes not mentioned herein.

- **Etcd nodes**

Nodes with the etcd role run etcd, which is a consistent and highly available key value store used as Kubernetes' backing store for all cluster data. The etcd database is consistently replicated between all etcd nodes

- **Controlplane nodes**

Nodes with the controlplane role run Kubernetes master components (excluding etcd).
A complete list of system requirements for SUSE RKE2 is part of the SUSE RKE2 documentation (refer to Rancher | RKE requirements).
For more detailed recommendations regarding downstream clusters architectures refer to:
SUSE Rancher | Recommended Cluster Architecture
Kubernetes | Creating Highly Available clusters with kubeadm
With increasing workload, the capacity of an SAP Data Intelligence cluster can be scaled by adding additional worker nodes and related resources like disc capacity. Having multiple SAP Data Intelligence systems serving different purposes and SAP Data Intelligence clusters extension by means of adding extra worker nodes may lead to a high number of physical nodes that will require maintenance, rack space, energy & cooling, etc. Different strategies enable the number of physical servers to be limited. Each potential deployment strategy has its specific advantages and limitations and may not be suitable for a specific SAP Data Intelligent system, e.g., a productive system.

The architecture of a Kubernetes cluster running SAP Data Intelligence application is a combination of SUSE Rancher RKE2 requirements, related to the basic Kubernetes architecture and requirements defined by SAP for running SAP Data Intelligence workloads.

- Controlplane and etcd must not share the physical nodes with worker nodes in production systems
- Controlplane and etcd may run on the same physical node, the master node

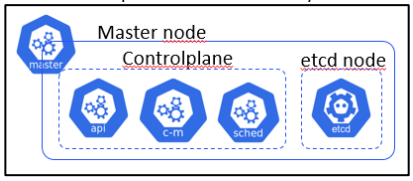


Figure 3: Master node

- For highly availability clusters a minimum of 3 nodes running etcd/controlplane are required
- Minimum nodes to run worker node role
  4 worker nodes for productive systems (SAP)
  3 worker nodes for non-productive systems (SAP)

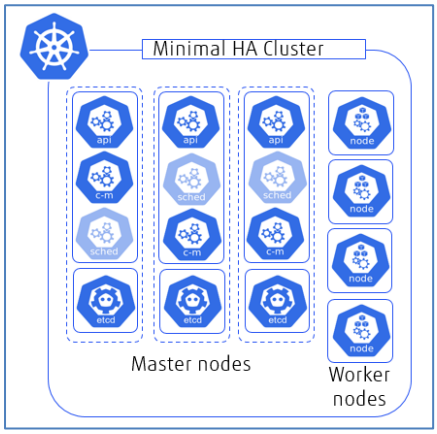


Figure 4: Minimal Kubernetes Cluster - Highly Availability

Standard deployment of a Kubernetes cluster running SAP Data Intelligence requires three physical servers running the controlplane and etcd services.This setup provides high availability since all roles (controlplane, etcd, worker) are redundant and is the recommended setup for all types of SAP Data Intelligence systems. Etcd and controlplane are critical for the stability of the cluster.

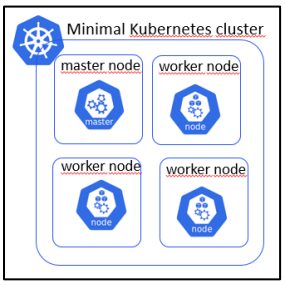


Figure 5: Minimal Kubernetes Cluster - no HA

When it comes to building Kubernetes clusters, single master node deployments provide a cost-effective solution. As such a cluster becomes unavailable in case of master node outage, this deployment method is not suitable for productive systems running core business processes or other types of systems, where guaranteed availability is required.Cluster node deployment on a virtualized server provides an additional possibility to decrease TCO. The drawbacks of the different deployment methods discussed above remain the same. If the Kubernetes cluster is deployed on physical nodes, where other Kubernetes clusters or other applications/databases (e.g., SAP HANA) are already running, it is necessary to ensure that the different systems can be isolated to a large extent to avoid negative effects on system behaviors in case of high load phases or error situations. At least one physical spare node and rolling upgrade procedures to minimize unnecessary planned downtimes are mandatory. Different growth rate and computing power requirements of different systems in combination with potentially different scaling possibilities (scale-out vs. scale-up) may lead to only suboptimal hardware use.

Even for big productive clusters a significant reduction of physical worker nodes can be achieved by deploying two virtual worker nodes on one server. However, if one worker node fails, the remaining worker nodes should have enough available capacity to run the services of the failed node.

Other possibilities of TCO reduction by sharing hardware between different systems are technically possible. Refer to relevant deployment guides and notes from the different system and virtualization vendors for a detailed discussion about various pros and cons of each solution. The links below provide a brief overview of the possibilities and restrictions of different deployment options with focus on SAP Data Intelligence, SAP HANA and VMware.

For more information refer to:
SAP | Installation Guide for SAP Data Intelligence on-premise - Planning your installation
SAP | Note 2856760 - Multiple SAP Data Hub installations in one Kubernetes cluster
SAP | Note 1681092 - Multiple SAP HANA systems (SIDs) on the same underlying server(s)
SAP | Note 2680623 - Multiple SAP HANA Scale-out Systems (SIDs) on the Same Underlying Servers

SAP | 2779240 - Workload-based sizing for virtualized environments
SAP | 2015392 - VMware recommendations for latency-sensitive SAP applications
SAP | 1492000 - General Support Statement for Virtual Environments
SAP | 2161991 - VMware vSphere configuration guidelines
SAP | SAP on VMware vSphere
SAP | SAP HANA Guidelines for running virtualized
SAP | Known Support Issues in Virtual Environments

SUSE Rancher| Installing Rancher in a vSphere Environment
SUSE Rancher | Best Practices for Rancher Managed vSphere Clusters

Cluster Management - SUSE Rancher Server Cluster

The majority of SUSE Rancher 2.x software runs on a Rancher Server, which includes all the software components used to manage the entire SUSE Rancher deployment.
The figure below illustrates a high-level architecture of SUSE Rancher 2.x. The figure depicts a SUSE Rancher Server installation that manages two downstream Kubernetes clusters. For the best performance and security, a cluster dedicated just for SUSE Rancher should always be present in production environments. Running user workloads on such a cluster is not advised. After deploying SUSE Rancher, you can create or import clusters that will run your workloads.
The figure below shows the dependencies of the SUSE Rancher Management cluster (single node and HA configuration) and the downstream aka user clusters based on RKE2 running your SAP Data Intelligence systems (e.g., DEV, QAS, Prod). The SUSE Rancher management server can manage both hosted (i.e., EKS, AKS, GKE) and IaaS based clusters (EC2, vSphere, Azure VM's, etc.) running other workloads as well. For more information refer to SUSE Rancher | Managed Kubernetes Cluster Operations.

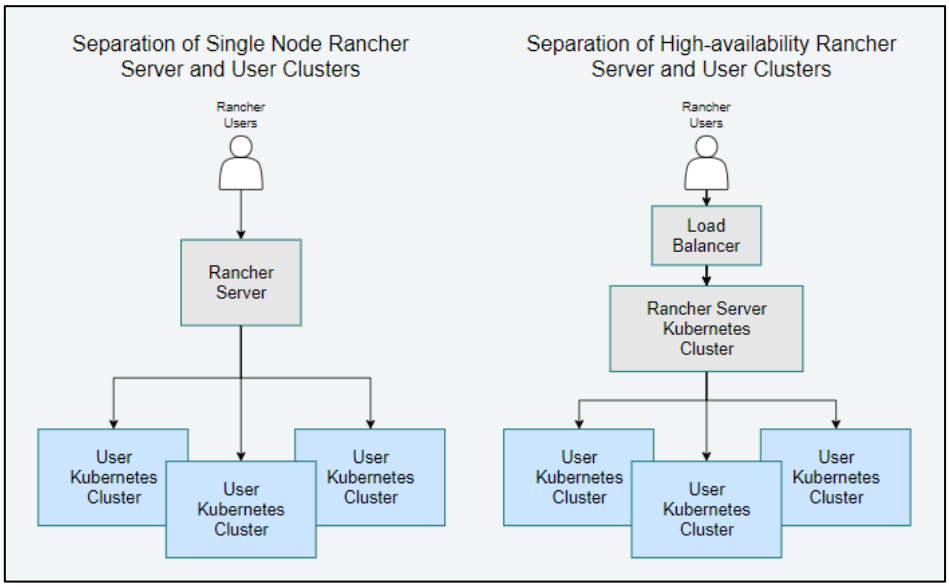


Figure 6: Seperation Rancher Management Cluster and User Clusters

The installation requirement for a SUSE Rancher Server cluster are mainly dependent on the number and size of managed user clusters:

- CPU and Memory

Hardware requirements are based on the size of your SUSE Rancher deployment. Provision each individual node according to the requirements. The requirements are different and depend on whether you are installing SUSE Rancher in a single container with Docker or you are installing SUSE Rancher on a Kubernetes cluster. For productive use deployment on a single docker cointainer is not sufficient due to missing redundancy.

- RKE2 and Hosted Kubernetes

These CPU and memory requirements apply to each instance with RKE2 installed. Minimum recommendations are outlined below:

DEPLOYMENT SIZE	CLUSTERS	NODES	VCPUS	RAM
Small	Up to 5	Up to 50	2	5 GB
Medium	Up to 15	Up to 200	3	9 GB

A detailed list of installation requirements can be found at SUSE Rancher | Hardware Requirements.

Network concept

Below is the proposal of minimum network requirements for SAP Data Intelligence systems:

- Administration network
- Datacenter internal network
  - Server network
  - Optional SAN (iSCSI) Storage network when using NetApp Trident
- User access network
- Internet for software updates, etc.

For sizing the required bandwidth of administration and user access networks, the vendor recommendations provide reliable information. Sizing of the server and storage network requirements depends on the usage scenarios and the data volumes expected as well as the storage technology used. Storage systems using data replication methods like SUSE Longhorn and NetApp StorageGrid may cause high network traffic by distributing redundant datafiles - blocks. The SAP Data Intelligence sizing guide provides methods for estimating needed network bandwidth and storage IO requirements in case of using Distributed Data Management.
(Explanation symbols used)

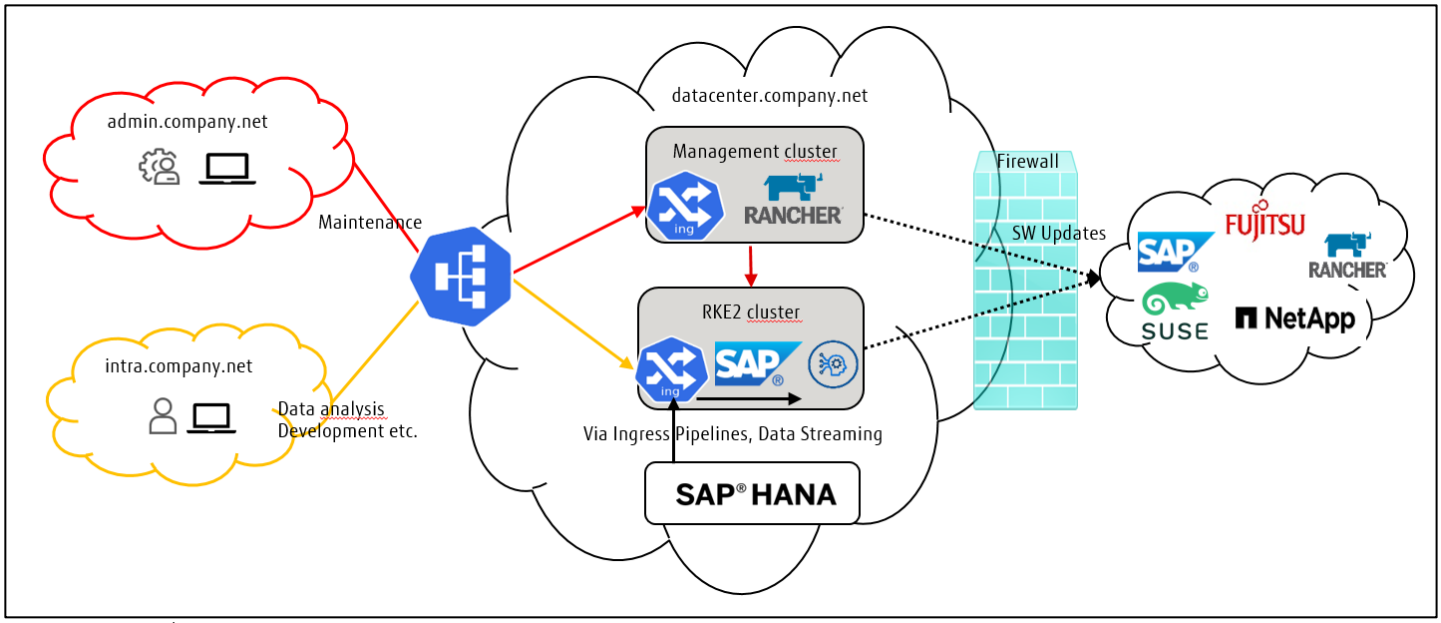


Figure 7: Netwok Integration Overview

Load balancing

Load balancing is an important concept for building highly available Kubernetes clusters and deploying application services. To ensure a transparent access to all applications a loadbalacing service provides a flexible and highly available entrypoint to the application. Load balancers create a gateway for external connections to access your cluster The load balancer directs the incoming requests across a pool of worker nodes, commonly referred to as backends. Since the load balancer presents itself as the endpoint for the site, the customers don't know anything about these backends. The load balancer tracks the health and number of connections to each backend, and it works according to its configured policy to evenly distribute the traffic. This scenario enables horizontal scaling, where a site can scale capacity by adding and removing backends.

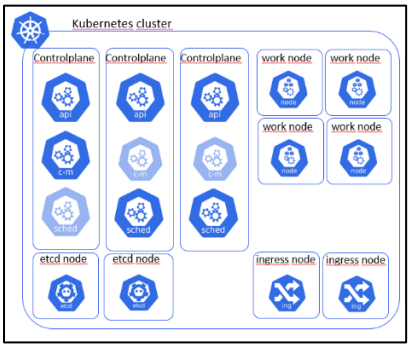


Figure 8: Deployment Ingress Nodes

SUSE Rancher supports two types of load balancers, namely:
SUSE Rancher | Layer 4 Loadbalancer
SUSE Rancher | Layer-7 Loadbalancer
An Ingress Controller is a daemon, deployed as a Kubernetes pod, that listens for requests to create or modify ingresses within the cluster and converts the rules in the manifests into configuration directives for a load balancing component. That component is either a software load balancer such as Nginx, HAProxy, or Traefik, or an external load balancer such as Amazon ALB or F5 B i g / IP.

For more information refer to:
SUSE Rancher | Architecture Recommendations
Kubernetes| services networking ingress

Step 3: Datacenter Infrastructure Layer - PRIMERGY Enterprise Server

FUJITSU Server PRIMERGY systems deliver workload-optimized x86 industry standard servers for any workload and business demand. Since there is no single server solution to meet all these needs, Fujitsu offers a broad server portfolio consisting of expandable tower servers, versatile rack-mount servers, density-optimized multi-node servers as well as GPU servers purpose-built for the demands of AI and VDI. While all these systems are designed to handle multiple workloads, each server is optimized for specific use cases. Whatever the size of your business - large enterprise with multiple sites, or a small or medium-sized company with limited space and budget - with the right choice of server, your IT can become the business enabler you have always wanted it to be (Fujitsu | Fujitsu PRIMERGY Server Models).

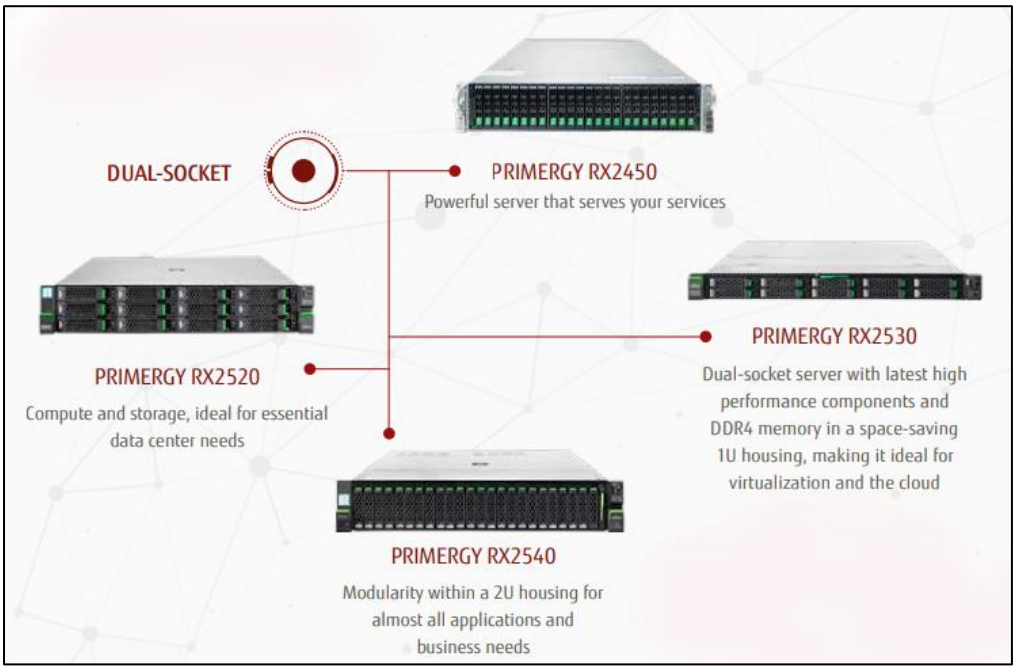


Figure 9: PRIMERGY Rack Server

FUJITSU Server PRIMERGY Rack Systems are versatile, rack-optimized servers that provide best-in-class performance and energy efficiency. You can rely on these servers as the backbone of your IT operations and easily scale them as you grow. Moreover, the PRIMERGY rack server portfolio is performance optimized for multi-application workloads to significantly increase the efficiency of IT operations and enable IT to respond faster to business needs of any size. Backed by more than 25 years of development and production at Fujitsu, these are extremely robust servers that you can rely on to take your business to the next level.

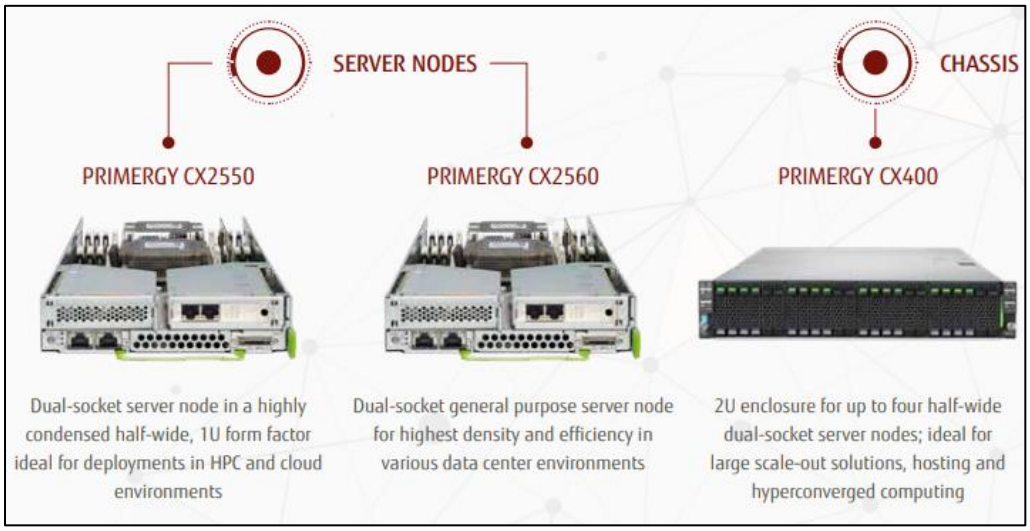


Figure 10: PRIMERGY Multinide Server

The PRIMERGY CX multi-node systems offer an ideal basis for cloud, hyper-converged and high-performance computing solutions. They provide data centers as well as branch offices with massive computing power while at the same time delivering best economics for server density, energy consumption, heat optimization and lower overall operating costs.

The selected PRIMERGY rack and scale-out servers complement each other perfectly to ensure that your SAP Data Intelligence landscape can be easily and quickly integrated into a wide variety of data center infrastructures and operational management concepts.

Cluster	Service Nodes	RX2520 M6	RX2530 M6	RX2540 M6	CX400 M6 CX2560 M5
RKE2 Downstream cluster	Worker	Ideal for high data volume configurations	Very good Ideal for virtualization, scale-out scenarios, and small databases as well as for high performance computing	Excellent expandability with up to 12x 3.5" SAS/SATA, up to 24x 2.5" SAS/SATA/NVMe, or the option to use up 64x EDSFF The server can be equipped with up to six NVIDIA GPU cards. Thus, the server also provides optimized performance for AI and HPC workloads	Very good high server density
	etcd	---	Very good	---	Very good high server density
	controlplane	---	Very good	---	Very good high server density
SUSE Rancher Server Management Cluster	Worker, etcd, controlplane	---	Very good	---	Very good high server density

The SAP Data Intelligence reference architecture consists of different components / services which handle special tasks and workloads. The table obove shows which of the servers mentioned is best suited for running SAP Data Intelligence services.

Step 4: Data Persistency Layer - Persistent Volumes & Object Storage

SAP Data Intelligence requires different kinds of persistent storage that should be considered during the installation planning process:

- Automatic provisioning of persistent volumes (RWO) in the Kubernetes Cluster
- A container registry for executing data pipelines
- An object storage for the purposes of Backup and Recovery, Machine learning scenarios and usage of streaming tables with SAP Vora (Distributed Data Management)

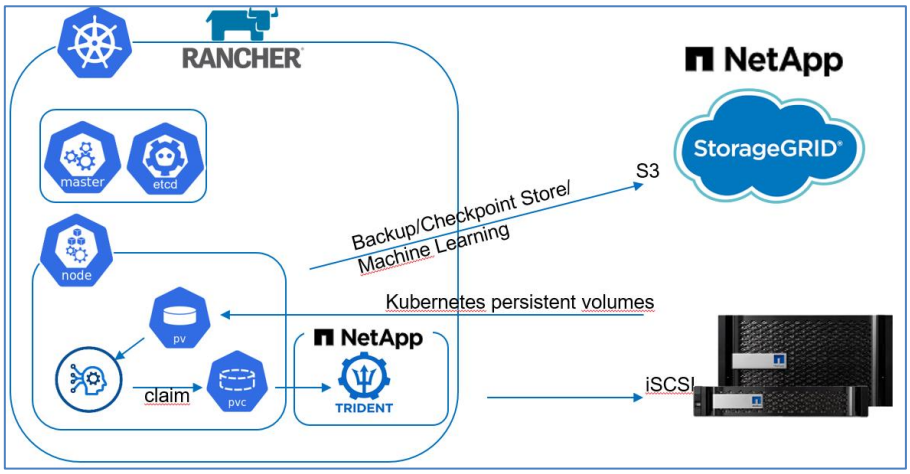


Figure 11: Storage Types

Currently (Q3 2021) SAP has realeased following products to build a reliant and performant storage layer for you SAP Data Intelligence systems For updates check SAP Note SAP | Note 2693555 - SAP Data Hub 2 and SAP Data Intelligence 3: Supported Remote Systems and Data Sources

Vendor	Kubernetes persistency volumes	Object Storage	Remarks
NetApp	Trident	StorageGrid	ok
SUSE	Longhorn	n/a	Not recommended as limited application functionality and missing backup/restore of your SAP Data Intelligence system
NetApp & SUSE	Longhorn	StorageGrid	ok

Data Persistency Layer – Kubernetes Persistency Volumes NetApp Trident for Kubernetes

Trident is a fully supported open-source project maintained by NetApp. It has been designed from the ground up to help you meet the sophisticated persistence demands of your containerized applications. Trident integrates natively with Kubernetes and its Persistent Volume framework to seamlessly provision and manage volumes from systems running any combination of NetApp's ONTAP (AFF/FAS/Select/Cloud), Element (HCI/SolidFire), or SANtricity (E/EF-Series) data management platforms.

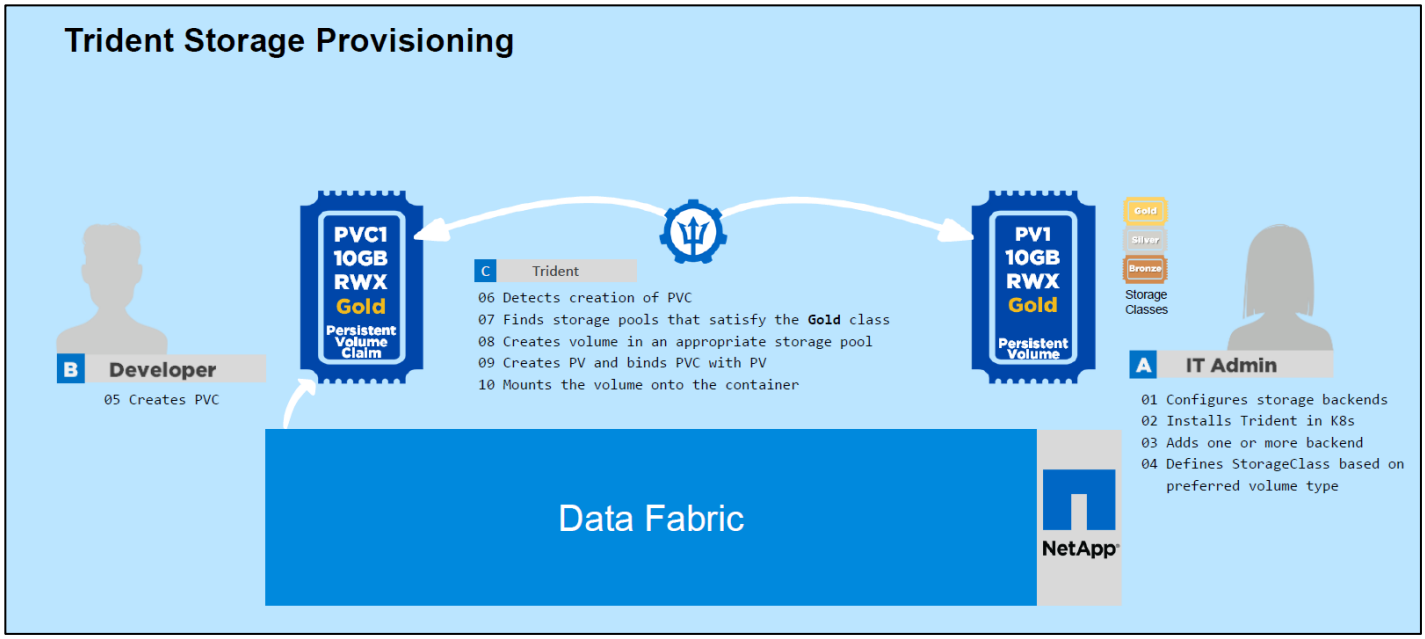


Figure 12: NetApp Trident Storage provisioning

Relative to other Kubernetes provisioners, Trident is novel in the following respects:

1. It is the first out-of-tree, out-of-process storage provisioner that works by watching events at the Kubernetes API Server, affording it levels of visibility and flexibility that cannot otherwise be achieved.
2. It is capable of orchestrating across multiple platforms at the same time through a unified interface. Rather than tying a request for a persistent volume to a particular system, Trident selects one from those it manages based on the higher-level qualities that the user is looking for in their volume.
3. Trident tightly integrates with Kubernetes to allow users to request and manage persistent volumes using native Kubernetes interfaces and constructs. It's designed to work in such a way that users can take advantage of the underlying capabilities of your storage infrastructure without having to know anything about it. Today, that infrastructure includes our **ONTAP** (AFF/FAS/Select/Cloud), **Element** (HCI/SolidFire), and **SANtricity** (E/EF-Series) data management software. That list continues to grow.

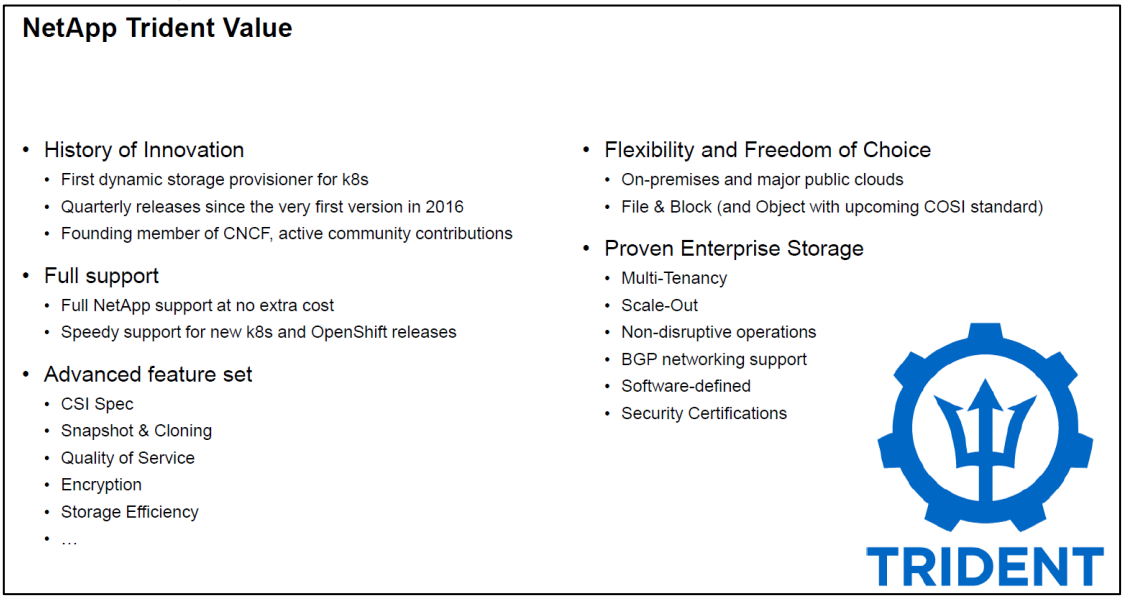


Figure 13: NetApp Trident values

For more information refer to:
NetApp | NetApp Trident Deployment & Configuration
NetApp | NetApp Trident for Kubernetes - Persistent Volumes
NetApp | Documentation on the possible parameters of a storage class
NetApp | Examples of different use cases of Trident

Data Persistency Layer – Kubernetes Persistent Volumes SUSE Rancher Longhorn

Alternatively, it is possible to substitute NetApp Trident by SUSE Rancher Longhorn. It is a lightweight, reliable, and powerful distributed block storage system for Kubernetes. Longhorn implements distributed block storage using containers and microservices as well as creates a dedicated storage controller for each block device volume and synchronously replicates the volume across multiple replicas stored on multiple nodes. The storage controller and replicas are orchestrated using Kubernetes.
The main features of Longhorn include:

- Enterprise-grade distributed block storage with no single point of failure
- Incremental snapshot of block storage
- Backup to secondary storage (NFS or S3-compatible object storage) built on efficient change block detection
- Recurring snapshots and backups
- Automated, non-disruptive upgrades (you can upgrade the entire Longhorn software stack without disrupting running storage volumes)
- An intuitive GUI dashboard

The figure to the below shows three instances with Longhorn volumes. Each volume has a dedicated controller, which is called the Longhorn Engine and runs as a Linux process. Each Longhorn volume has two replicas, and each replica is a Linux process. The arrows in the figure indicate the read/write data flow between the volume, controller instance, replica instances, and disks. By creating a separate Longhorn Engine for each volume, the function of other volumes is not impacted if one controller fails.
For more information refer to:
Longhorn | What is Longhorn?
Rancher | Longhorn - Persistent Volumes

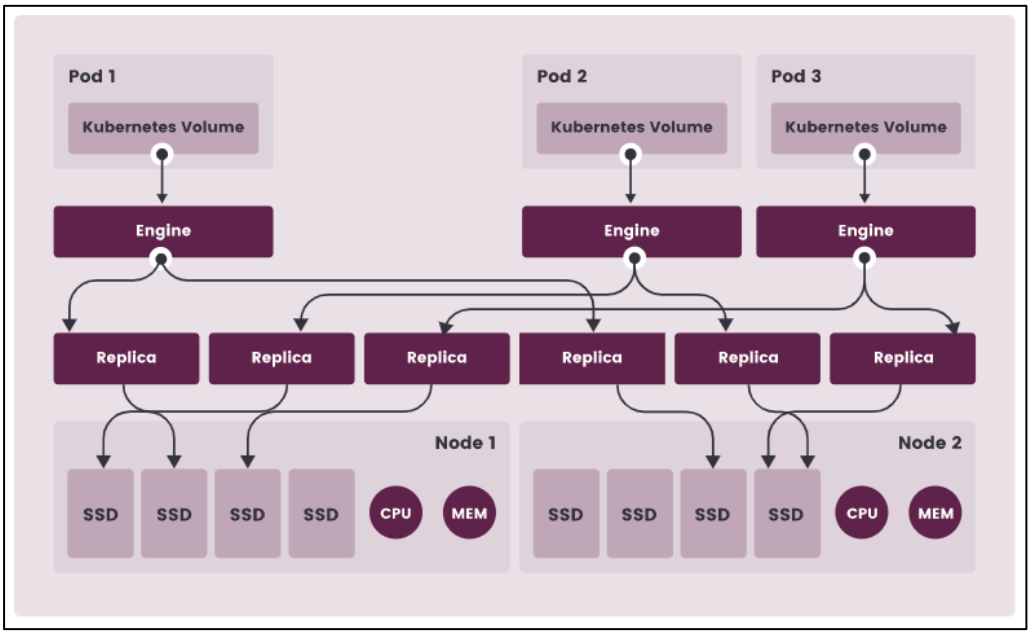
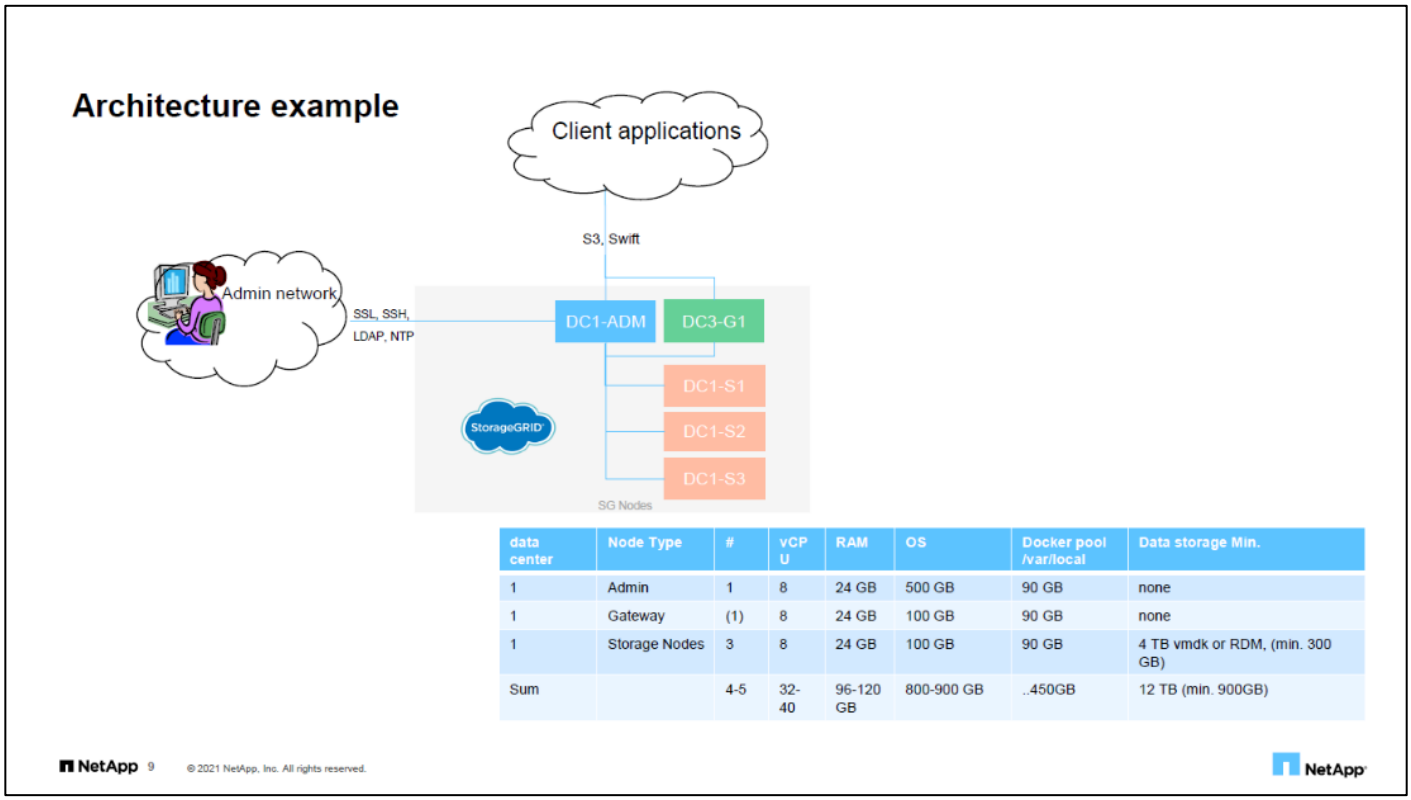


Figure 14: SUSE Longhorn Architecture

Data Persistency Layer – Object Storage NetApp StorageGrid

NetApp® StorageGRID® is a software-defined, object-based storage solution that supports industry standard object APIs like the Amazon Simple Storage Service (S3) API. It allows you to build a single name space across up to 16 data centers worldwide, with customizable service levels for metadata-driven object lifecycle policies. The integrated lifecycle management policies optimize where your data lives throughout its lifecycle. StorageGRID optimizes your data durability and availability across multiple geographies. It enables hybrid cloud workflows - whether your data is on premises or in a public cloud - to fit your business demands with access to Amazon Simple Notification Service (SNS), Microsoft Azure Blob, Amazon Glacier, Elasticsearch, and similar services. StorageGRID software-defined object storage suite supports a wide range of use cases across public, private, and in hybrid multi-cloud environments seamlessly. With industry leading innovations, NetApp StorageGRID stores, secures, protect, and preserves unstructured data for multi-purpose use including automated lifecycle management for long periods of time. For more information refer to:
NetApp | NetApp StorageGRID - Object Storage
NetApp | Introduction to StorageGrid
NetApp | Deploying StorageGRID in a Kubernetes cluster



data center	Node Type	#	vCPU	RAM	OS	Docker pool /var/local	Data storage Min.
1	Admin	1	8	24 GB	500 GB	90 GB	none
1	Gateway	(1)	8	24 GB	100 GB	90 GB	none
1	Storage Nodes	3	8	24 GB	100 GB	90 GB	4 TB vmdk or RDM, (min. 300 GB)
Sum		4-5	32-40	96-120 GB	800-900 GB	..450GB	12 TB (min. 900GB)

Figure 15: NetApp StorageGRID Architecture

NetApp StorageGRID is available as a software-defined deployment across heterogeneous hardware. Deployment is possible on:

- Software-defined nodes using virtual machines running in VMware
- Software-defined nodes using Docker containers running in third-party operating systems

A basic NetApp StorageGRID requires at minimum

- 3 Storage nodes
- 1 Admin node
- 1 optional Gateway node
- Load balancers, StorageGRID provided or 3rd party

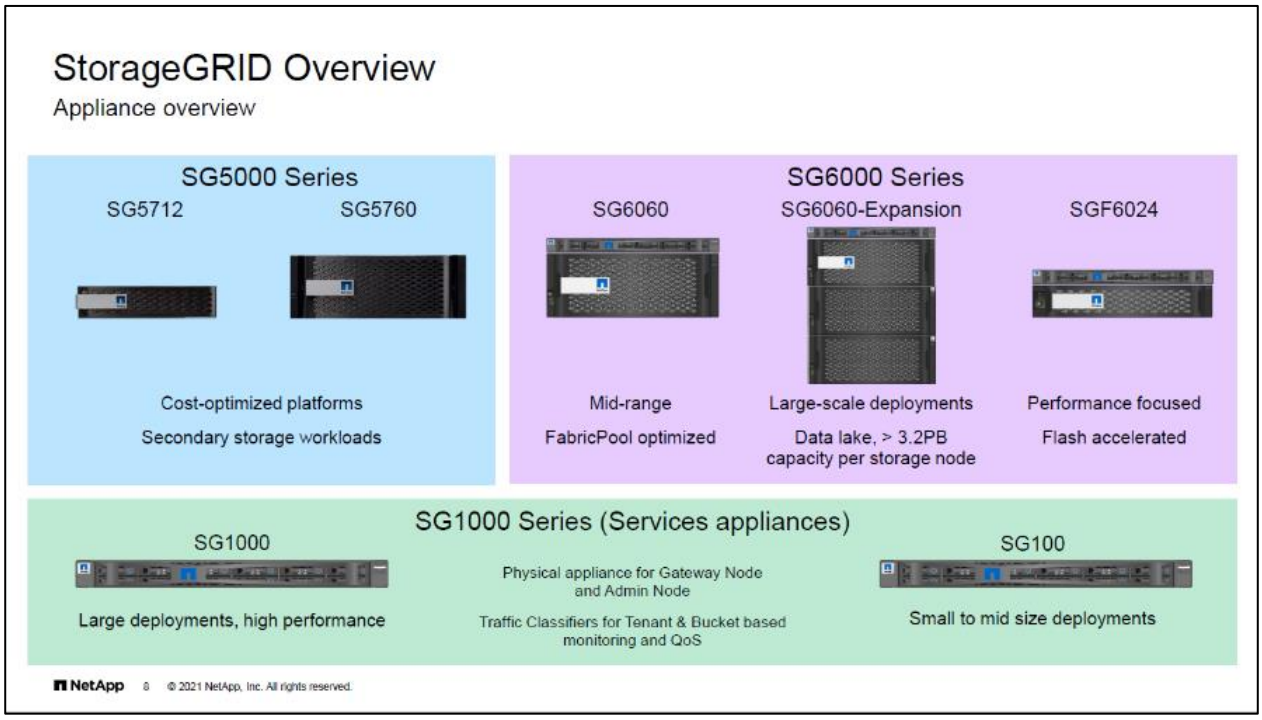


Figure 16: NetApp StorageGRID Appliance Overview

Alternatevely engineered hardware appliances built on NetApp E-Series can also been used. Technical details are described in the NetApp StorageGRID datasheet: [NetApp | NetApp StorageGRID data sheet]

Advantages of the StorageGRID system include:

- Massively scalable and easy-to-use global data repository for unstructured data
- Standard object storage protocols:
  - Amazon Web Services Simple Storage Service (AWS S3)
  - OpenStack Swift
- File system storage protocols (requires StorageGRID NAS Bridge virtual appliance):
  - Network File System (NFS)
  - Server Message Block (SMB)
- Hybrid cloud enabled. Policy-based management stores objects to public clouds, such as Amazon Web Services (AWS). StorageGRID platform services enable content replication, event notification, and metadata searching on public clouds.
- Flexible data protection to ensure durability and availability. Data can be protected using replication and layered erasure coding. At-rest and in-flight data verification ensures integrity for long-term retention. Tape and S3-compatible public cloud are available as an integrated storage tier
- Dynamic data lifecycle management policy to control storage cost. Administrators can create rules that manage data lifecycle at the object level, and customize data locality, durability, performance, cost and retention time
- Compliance with regulatory requirements. Designed to meet the requirements of the Securities and Exchange Commission (SEC) Rule 17a-4(f)
- Non-disruptive upgrade and operations. Maintain access to content during upgrades, expansion, decommission, and maintenance procedures
- Federated identity management. Integrates with Active Directory, OpenLDAP, or Oracle Directory Service for user authentication. Supports single sign-on (SSO) using the Security Assertion Markup Language 2.0 (SAML 2.0) standard to exchange authentication and authorization data between StorageGRID and Active Directory Federation Services (AD FS)

Storage requirements are low if only Data Governance and Data Pipelines are used. Using Distributed Data Management and Machine Learning, the required data volume size, IOPS and network bandwidth have to be carefully sized at least during initial sizing as documented in the sizing guide. The storage system architecture and the type of data which have to be processed also play an important role. Storage systems that use replication methods to achieve data protection may cause a significant network and data volume overhead, especially if data compression rate is low.

SAP Data Intelligence functional areas	Component	Kubernetes Persistent volumes (GB)	Checkpoint store (GB)	Ephemeral disks (GB)	Docker registry (GB)
Static components Data Governance Data Pipelines		240 GB		100 GB per worker node (Local disks)	90 GB/Kubernetes cluster (SAP Data Intelligence system)
Distributed Data Management	Relative disk engine data volumes <= 5 TB	= Source Data Footprint / Compression Factor + Max Load Data Size / Compression Factor + Overhead	Disk (GB) = 3.3 * Streaming Data Footprint / Compression Factor		
Distributed Data Management	Relative disk engine data volumes > 5 TB	Number of Disk Engine Instances = MAX (CEILING (Source Data Footprint / 10,000 GB);3) Number of Disk Engine Instances * (Source Data Footprint / Compression Factor * Number of Disk Engine Instances) + 100 GB / (Compression Factor * Number of Disk Engine Instances) + 100 GB)			
Distributed Data Management	Distributed log	Streaming tables are used, and checkpoint store is enabled: ========== Disk (in GB) = 0.0001 GB/MB * Ingestion Rate * 3,600s Streaming tables are used, and checkpoint store is disabled: ========== Disk (in GB) = Streaming Data Footprint * Overhead Factor			
Miscellaneous Components					
Machine Learning					
Backup/Restore					

For additional explanations and procedures related to sizing of networks and disc IOPS, refer to the most current version of SAP | SAP Data Intelligence Sizing Guide Appendix B .
For a detailed overview of the storage requirement, refer to the most current version of SAP | SAP Data Intelligence - Installation Guide V. 3.14

Despite an overwhelming number of technical features for every storage product, the different technical characteristics mentioned in the table are of high importance as they are affecting the needed server, network and storage configuration substantially. This makes it difficult to have a simple set of rules for such a type of configuration. At this stage of the project, it is recommended to use the metrics to calculate the needed data volumes for the relational disk engine and the checkpoint store as well as the calculation formulas for the required network bandwidth. Depending on your preferences for a specific type of storage, a project-based sizing approach for your storage environment should be implemented.

Summary - Bringing bits and pieces together

RKE2 downstream cluster

Depending on the importance of the SAP Data Intelligence System for your core business processes, each cluster should be treated as productive or non-productive. The number of cluster management nodes increases with the need for high availability of the system. High availability is provided by node redundancy - an odd number of management nodes (higher than 3) is required at all times. For increased high availability requirements, additional deployments that require more than 3 management nodes are possible. This aspect is not covered herein. Project-based consulting needs to be performed if required. This approach applies also to clusters stretched over multiple datacenters or geographical regions.The number of worker nodes needed depends on your system type (DEV, QAS, TRN, PRD) and the used business scenarios, and is determined by proper sizing.

SUSE Rancher Management cluster

Despite the fact that all 3 roles (worker, etcd, controlplane) are employed on each SUSE Rancher Management cluster node, high availability for a SUSE Rancher Management cluster comes with node redundancy as well. An odd number of SUSE Rancher Management nodes equal or higher than 3 is required at all times.
Even one SUSE Rancher Management cluster is capable of managing all your productive and non-productive clusters; however, it is recommended to have a minimum of one productive SUSE Rancher Management cluster that manages your productive downstream cluster, and additionally one non-productive SUSE Rancher Management cluster for training and testing purposes for systems management personnel as well as for non-productive downstream cluster management. More complex distributed deployments are possible but require additional project-related consulting.

System types (Staging, PRD)	RKE2 Downstream cluster	SUSE Rancher Management Cluster
Min. number of clusters	Number of required SAP Data Intelligence systems	1
Min. number of worker nodes	4	3 (high availability)
Min. number of master nodes (etcd & controlplane)	3 (high availability)	
Kubernetes data persistency, mandatory	NetApp Trident alternatively SUSE Rancher Longhorn 250 GB/cluster not using Vora database For Vora scenarios appropriate sizing is mandatory	
Object Store; required if Backup/restore Machine Learning Checkpoint storage is used	NetApp StorageGrid; appropriate sizing is mandatory	

System types (DEV, QAS, TRN)	RKE2 Downstream cluster	SUSE Rancher Management Cluster
Min. number of clusters	Number of required SAP Data Intelligence systems	1
Min. number of worker nodes	3	1 (no high availability) 3 (high availability)
Min. number of master nodes (etcd & controlplane)	1 (no high availability) 3 (high availability)	
Kubernetes data persistency, mandatory	NetApp Trident alternatively SUSE Rancher Longhorn 250 GB/cluster not using Vora database For Vora scenarios appropriate sizing is mandatory	
Object Store; required if Backup/restore Machine Learning Checkpoint storage is used	NetApp StorageGrid; appropriate sizing is mandatory	

Other aspects that need to be considered

This document gives a rough impression what an already planned landscape can look like, how many nodes are required and which Fujitsu PRIMERGY server type fits best to different node types. For a complete system configuration additional points should be decided.

RKE2 Cluster			
Worker nodes	**Component**	**Dependent on**	**Additional information**
SAP Data Intelligence worker node size	32 GB RAM/8 CPU 64 GB RAM/16 CPU	Dependent on usage scenarios	Pipelines & Workflow; Data Governance Machine Learning & Distributed Data Management data volume > 5 TB
CPU/RAM	Intel XEON Silver/Gold	Workload dependent	CPU number=number of CPU req. by Data Intelligence + 2 CPU RAM size=number of RAM req. by Data Intelligence + 2 GB
Nics type and number/node	Redundant configuration Protocol iSCSI	Bandwidth dependent on usage scenarios storage layer	ML, VORA Ingestion rates Connected systems NetApp Trident & StorageGrid
Internal disks	2*SSD RAID1 Optional: Additional SSD for Kubernetes persistent volumes if SUSE Rancher Longhorn is used	SSD size dependent on usage scenarios/data volume. For Non-Vora scenarios 250 GB persistent volume size per cluster	Operating system Longhorn req. internal SSD for Kubernetes persistent volumes
External storage Kubernetes persistent volumes S3 compatible object storage	Persistent volumes provided by NetApp Trident Object store volumes provided by NetApp StorageGrid	Data volume sizing & usage scenarios	Kubernetes persistent volumes are mandatory and can be realized by SUSE Rancher Longhorn or NetApp Trident Backup/Restore Checkpoint store (Vora) Machine Learning
GPUs		Depends on machine learning req.	
OS	SLES		
PRIMERGY server type	Rackserver Multi-node server	Standard worker node Machine learning GPU worker node Standard worker node High density	PG RX2530 M6 PG RX2540 M6 PG CX400 M6 & PG CX2560 M6
Other	Licenses; subscriptions	Products	

RKE2 Cluster			
Master nodes (etcd & controlplane co-located)	Component	Dependent on	Additional information
CPU/RAM	Intel XEON Silver/Gold	Number of clients, requests per second & Data volume	Min. 2 to 4 vCPU Min 2 to 4 GB RAM
NICs		Number of clients, requests per second & Data volume	Depends on workload sizing
Internal disks	SSD		2*SSD RAID1 (OS) Additional SSDs for etcd db
OS	SLES		
PRIMERGY server type	Rackserver Multi-node server	Master node Master node High density	PG RX2530 M6 PG CX400 M6 & PG CX2560 M6
Other	Licenses; subscriptions	Products	

For a complete overview, refer to: etcd | etcd hardware requirements.

SUSE Rancher Management Cluster			
SUSE Rancher Manage-ment cluster nodes	Component	Dependent on	Additional information
CPU/RAM NICs		CPU and memory requirements apply to each instance with RKE2 installed. Minimum recommendations are outlined here	Rancher | Installation Requirements
Internal disks	2*SSD RAID1 Optional: Additional SSD for Kubernetes persistent volumes if SUSE Rancher Longhorn is used	SSD size dependent on usage scenarios/data volume. For Non-Vora scenarios 250 GB persistent volume size per cluster	Operating system Longhorn req. internal SSD for Kubernetes persistent volumes
External storage Kubernetes persistent volumes S3 compatible object storage	Persistent volumes provided by NetApp Trident Object store volumes provided by NetApp StorageGrid	SUSE Rancher Server hardware recommendations	Kubernetes persistent volumes are mandatory and can be realized by SUSE Rancher Longhorn or NetApp Trident Backup/Restore
OS	SLES		
PRIMERGY server type	Rackserver Multi-node server	Master node Master node High density	PG RX2530 M6 PG CX400 M6 & PG CX2560 M6
Other	Licenses; subscriptions	Products	

FUJITSU SAP Data Intelligence Services & Use Scenarios

SAP Data Intelligence can be added to the overall architecture in different functions. Depending on the Use Cases and demand, SAP Data Intelligence can support managing company data company in various ways. The main difference with reference to other systems or solutions is combining all data from various sources into the data management layer, while keeping the data in the source. The main purpose is to work as the data management layer, which supports data integration, orchestration and governance.

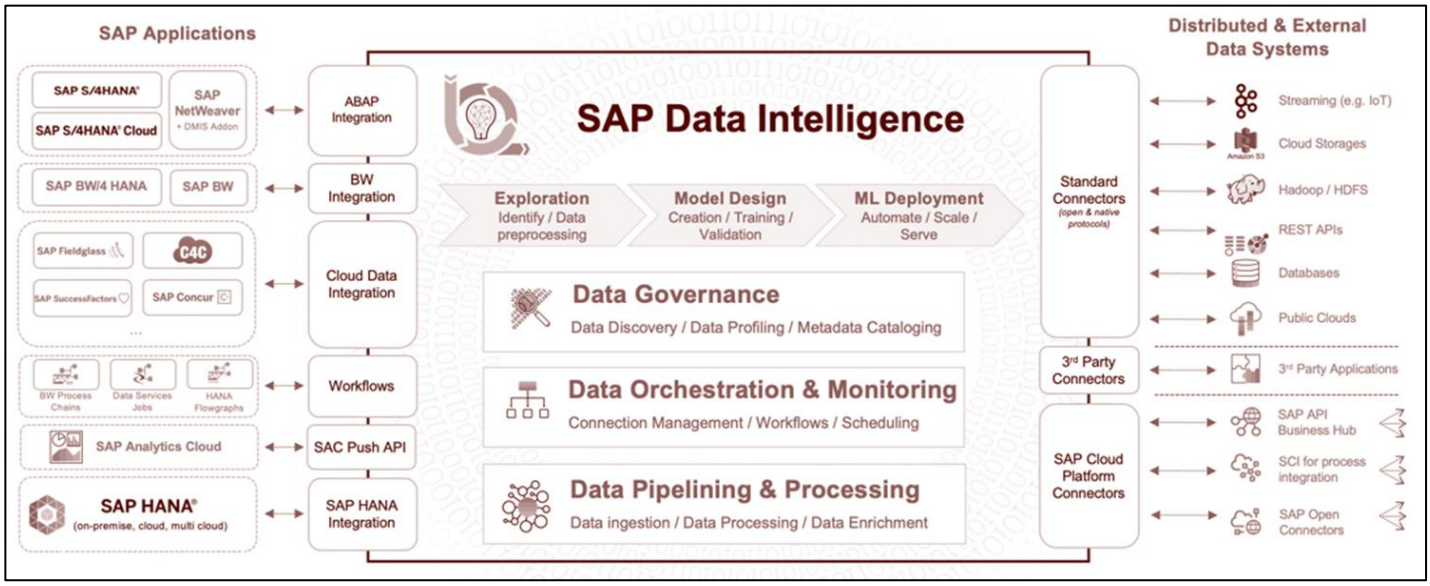


Figure 17: SAP Data Intelligence Functional Modules

Data integration is covered by a variety of interfaces, which includes SAP-specific APIs as well as standard or Third-Party APIs connecting non-SAP applications or sources. All processes can be handled and monitored by SAP Data Intelligence with pipelines, which can be created, edited and managed within the build in the modeler. The modeler is a user-friendly interface, which uses drag&drop for improved user experience. After including relevant systems, workflows can be set up using operators, which supports the user experience when building up processes within SAP Data Intelligence. Data can be harmonized, transformed and enriched before being moved to another system, application or machine learning models.

For data orchestration & monitoring SAP Data Intelligence delivers connection management, workflows and scheduling. Relevant system connections can be created using predefined masks, where for instance system IDs and users need to be entered. Data from connected systems can be managed with workflows. These workflows can be scheduled or proceeded directly. Workflows can be monitored, and the results analyzed. For complex analysis additional services can be included.

It is possible to use SAP Data Intelligence as a central instance to manage processes and authorizations from different systems. This leads to optimized compliance and data governance of the IT within your company. Functions for intelligent data correlation from new sources are available. Data relationships can be displayed in a metadata catalogue. Validation functions, which can be defined, support high data quality maintenance.

SAP Data Intelligence enables direct Machine Learning development. Common statistical libraries, for example R Libraries, are used to create, design, and train models. These libraries can be included in SAP Data Intelligence and used in workflows. Fujitsu offers various examples, e.g., anomaly detection with SAP Data Intelligence or automated purchasing process use cases, where Machine Learning models are used and implemented by application consultants and data scientists.

Furthermore, Fujitsu SAP Data Intelligence services cover data driven customer journeys from use case scoping to detailed implementation. Fujitsu has resources that enable all the mentioned topics to be consulted on both the conceptual and technical level. Anomaly detection, predictive maintenance or quality assurance are just a few examples Fujitsu has already implemented for their own factory, which also includes FUJITSU Edge Computing.

FUJITSU Manufacturing Maintenance Purchasing Use Case

The main aim of the Fujitsu Manufacturing Maintenance Purchasing Use Case is a central-owned automated re-order process in manufacturing environments.

Use case on the business level

SAP Analytics Cloud dashboards help a production manager viewing a high-level dashboard on a mobile device of one Fujitsu production hall and 4 product lines. In case a product line gets a red alert, the production manager can drill down on his device to product line and machine level. This red alert is triggered by a possible production outage prediction for future. The corresponding part within the production line gets highlighted and the production manager gets a notification on his device, while doing other tasks in his production hall. The order was automatically prepared, and the production manager pushes a button on his device to release the workflow of ordering the suggested parts avoiding a production breakdown in the next days. Afterwards an email is sent out to the purchasing manager, who will receive a link to confirm the prepared order in SAP Hybris. As soon as the order is confirmed, the color in the dashboards for the specific parts gets yellow and the production manager can see the purchasing and delivery date within the dashboard.

FUJITSU Manufacturing Maintenance Purchasing Use Case at the business level

Use case	Automated maintenance purchasing process
Use case description	Central-owned automated re-order process in manufacturing environments
Actor	Production Manager, Maintenance Manager, Purchasing Manager, (Finance Manager)
Pre-conditions	Maintenance contracts between machine supplier and owner exist Issue classifications and Machine Learning models have to be implemented
Basic Flow	1. Risk of machine outage is identified in Run Time Optimization Dashboard 2. Information about risk is pushed to Production and Maintenance Managers 3. Affected spare part is classified 4. Delivery company of spare part is known 5. System prepares order in shop 6. Notification via email to Purchasing Manager is sent out 7. Purchasing Manager clicks on link in email and confirms prepared order in shopping card 8. Order and Delivery dates are created and confirmed 9. Notification to Production and Maintenance Manager about Order and Delivery dates
Post-conditions	Order with delivery dates is shown in Run Time Optimization Dashboard

Use case at the technical level

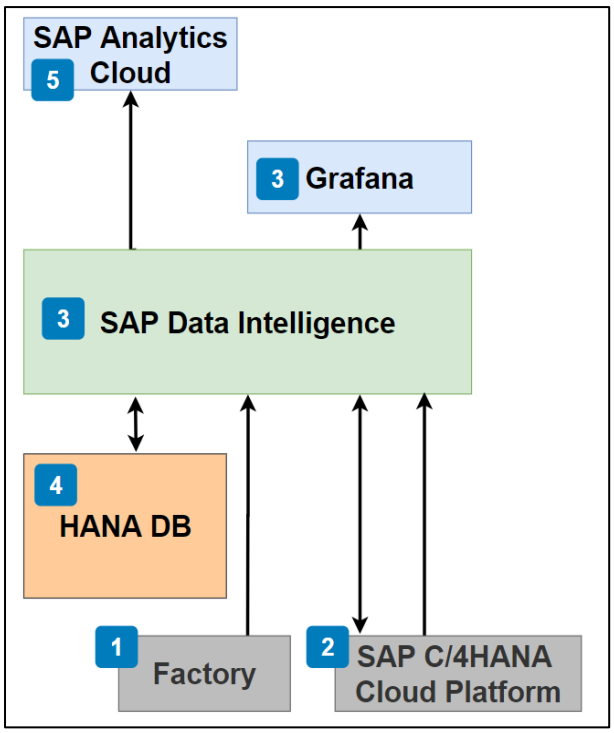


Figure 18: FUJITSU Manufacturing Maintenance Purchasing Use Case

SAP Analytics Cloud and Grafana are used as Front Ends to show dashboards of production lines. While SAP Analytics Cloud delivers high-level dashboards, Grafana covers a more detailed view with time series information where you can find the current stream data per second on a timeline. INTELLIEDGE is the Interface we use to get data from OT into IT landscape and send sensor data from production lines to SAP Data Intelligence. Data is orchestrated by SAP Data Intelligence and saved in HANA DB. Incoming data is analyzed by predictive Machine Learning models within SAP Data Intelligence, whether or not an anomaly is detected. A prediction is made for the next orders according to the classifications the models are trained with. Data is saved and updated in HANA DB. If data is older than 2 months, it is stored in cheap storage (Hadoop). If an anomaly is detected in SAP Data Intelligence, information will be sent to SAP Hybris and an order will be prepared. Information about the prepared order is sent to SAP Data Intelligence. The specific part is highlighted within SAP Analytics Cloud dashboards and an alert appears to show information about the outage of a machine, a broken part and order information. SAP Data Intelligence sends an email to the purchasing manager for information. After the workflow is successfully released by the production manager, the purchasing manager clicks on a link to go to the prepared shopping cart in SAP Hybris. After ordering a replacement for the broken part, data is sent back to SAP Data Intelligence with purchasing and delivery dates. SAP Data Intelligence collects information from SAP Hybris and updates HANA DB with the purchasing and delivery dates. Dashboards in SAP Analytics Cloud and Grafana show a hint that the broken parts within the production line are purchased and will be delivered.

FUJITSU Edge Computing

With the emergence of 5G connectivity, intelligent devices and Industry 4.0, Operational Technology (OT) and traditional IT are set to converge into Industrial IoT. Fujitsu is committed to bridging the OT/IT gap that enables businesses to face their unique challenges with the right solutions and helping them start their Edge Computing journey. Businesses need to understand and harness the advantages that Edge Computing provides when it comes to handling massive data sets generated by industrial systems and appliances at the edge of their enterprise. Applying deep analytics to such data enables businesses to gain and apply insights to further improve operational efficiencies in the long run.

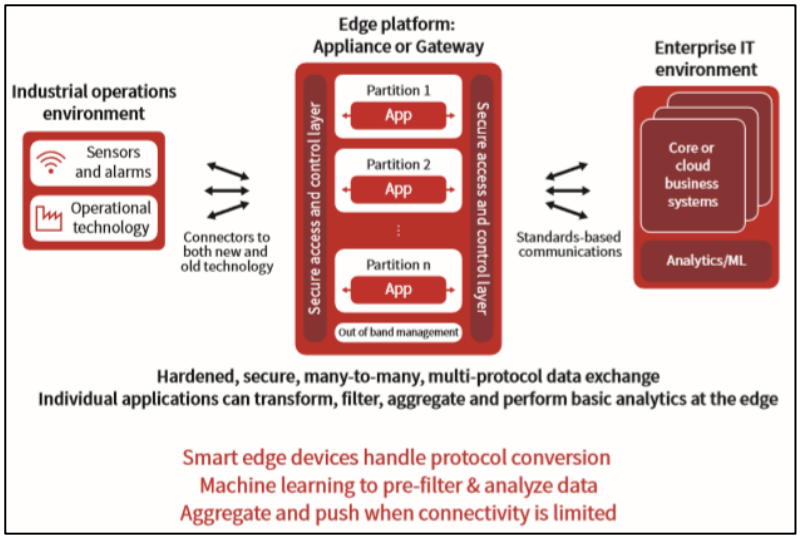


Figure 19: FUJITSU Edge Computing

The FUJITSU Manufacturing IoT Solution INTELLIEDGE™ Appliance and Gateway systems offer a scalable platform for a broad range of edge computing applications and scenarios in the Industry 4.0 and IIoT context. This typically ranges from simple gateway functionalities to fog computing scenarios that employ Artificial Intelligence (AI) and Machine Learning (ML) to make use of the data that gets generated at the edge of the enterprise.

INTELLIEDGE Appliance sits at the nexus between your business and your marketplace, and manages the collection, control and actuation of the vast amounts of data that are being generated by sensors in operational environments. INTELLIEDGE Gateway is a flexible and configurable hardware offering with pre-installed operating system that can be integrated easily into your enterprise Industry 4.0, IoT and IIoT initiatives. For more information refer to:

https://www.fujitsu.com/de/products/computing/pc/edge-computing/

FUJITSU INTELLIEDGE Appliance A700

Figure 20: FUJITSU INTELLIEDGE Appliance/Gateway rear side

FUJITSU IoT Solution INTELLIEDGE™ A700 Appliance combines the right mix of hardware components and software solutions helping organizations harness valuable operation data that keeps getting generated at the edge of their enterprise. Emerging technologies - such as 5G connectivity - coupled with interconnected and intelligent devices have already brought about a convergence of Operational Technology (OT) and traditional IT leading to the rise of Industrial Internet of Things (IIoT). INTELLIEDGE A700 is a converged IoT appliance ready to apply IoT applications for organizations that are looking into optimizing and improving operational efficiency by utilizing AI, Deep Analytics and Machine Learning with an intention to gain better insights into their business. For more information refer to:
https://www.fujitsu.com/de/products/computing/pc/edge-computing/intelliedge-a700-appliance/

Figure 21: FUJITSU INTELLIEDGE Appliance/Gateway front side

FUJITSU INTELLIEDGE Gateway G700

FUJITSU IoT Solution INTELLIEDGE™ G700 Gateway enables organizations to harness the data that gets generated at the edge of their enterprise. The Gateway device offers best-in-class industrial-grade hardware designed to operate in challenging environments to acquire, analyze and process data for better and deeper insights into the enterprise. It is a ready-to-run solution for organizations that want to deploy quickly without delay and start their Industry 4.0 journey. INTELLIEDGE G700 Gateway is a future-ready and easily customizable edge computing device that can be expanded and configured to match and scale to every requirement in an industrial IoT (IIoT) context. For more information refer to:
https://www.fujitsu.com/de/products/computing/pc/edge-computing/intelliedge-g700-gateway/

FUJITSU Integrated System PRIMEFLEX

Building data center infrastructures is increasingly complex, error-prone, time-consuming, risky and expensive. FUJITSU Integrated System PRIMEFLEX reduces complexity and risk, shortens time to value and reduces cost.
PRIMEFLEX is a pre-defined, integrated and tested combination of servers, storage, network connectivity and software. Fujitsu's PRIMEFLEX integrated system family includes both classical converged and hyper-converged systems. It comprises ready-to-run factory-installed solutions and reference architectures which are easily adjusted to customer-specific requirements. Along with its reference architectures, Fujitsu makes detailed configuration and installation descriptions available as a standard. Optional deployment and integration services ensure a smooth integration into the on-site environment. To ensure simplified operation and maintenance, Fujitsu provides support at solution level together with further data center services, including managed and hosting services. For more information refer to:
https://www.fujitsu.com/global/products/computing/integrated-systems/

Offerings for SAP Environments

A digital transformation requires platform solutions that support consistent and uniform management of entire SAP landscapes, including S/4HANA and SAP HANA, for every kind of IT provisioning model. As a trusted SAP Global Partner for almost 50 years, Fujitsu has successfully helped thousands of customers worldwide to simplify, innovate and grow. Certified hyper-converged infrastructure offerings for SAP environments allow running one infrastructure solution for all types of applications in-line with demands of the digital era. Furthermore, Fujitsu SAP SystemInspection Service - designed as a consulting package - helps customers to fully assess their SAP environment with an in-depth performance overview to best prepare for SAP S/4HANA migration whilst optimizing TCO. Turn your PRIMEFLEX for SAP infrastructure into a hybrid cloud with Fujitsu Software Enterprise Service Catalog Manager. For more information refer to:
https://www.fujitsu.com/global/products/computing/integrated-systems/sap/index.html

PRIMEFLEX for SAP HANA

Fujitsu's SAP HANA expertise, infrastructures and services enable customers to fully exploit the potential of the SAP HANA platform. Based on Fujitsu Integrated System PRIMEFLEX for SAP HANA optimized for in-memory technology, vast amounts of data (Big Data) can be analyzed in real time, safely and securely, either on premise or in the cloud so as to enable smarter business decisions based on greater business insight. Applications are also accelerated, enhancing business processes and enabling organizations to become true real-time enterprises.

PRIMEFLEX for SAP HANA is backed by 40 years of experience in delivering fast, secure, high availability implementations with optimized TCO, successfully reducing complexity. The pre-defined and pre-tested infrastructure solution is based on SAP certified components and leverages technologies of our specialist partners, such as NetApp, VMware and SUSE. It covers everything, from pre-installed scale-up systems and VMware virtualized platforms as well as individual scale-up and scale-out concepts in line with the SAP HANA Tailored Data Center Integration (TDI) approach right through to customized disaster-tolerant set-ups.

PRIMEFLEX for SAP HANA is supplemented by additional services for all project phases; from decision-making and financing up to ongoing operations. For more information refer to:
https://www.fujitsu.com/global/products/computing/integrated-systems/sap-hana.html

PRIMEFLEX for SAP Landscapes

Fujitsu's PRIMEFLEX for SAP Landscapes enables a simplified and secure set-up of infrastructures optimized for SAP applications and databases. The pre-defined, pre-integrated and pre-tested combination of servers, storage, network connectivity and software from Fujitsu and specialist partners, such as NetApp, VMware and SUSE, ensure highest quality and fast time to value.

Powered by the FlexFrame Orchestrator software component, PRIMEFLEX for SAP Landscapes facilitates the operation of SAP applications, SAP databases and the SAP HANA platform and makes it faster and more effective. It simplifies the management of complex SAP environments, optimizes planning, operation and change management and reduces costs by up to 90% whilst increasing agility by up to 50%.

Comprising very advanced orchestration and administration capabilities, PRIMEFLEX for SAP Landscapes provides an optimized operational concept for the entire SAP landscape. Centralized SAP software components can be dynamically deployed across physical and virtual resources based on business demand. The result is faster provisioning of innovations resulting in an empowered and agile business. This can be applied for all IT provisioning models, on premise, as a managed or hosting service, or deployed in the cloud. For more information refer to:
https://www.fujitsu.com/global/products/computing/integrated-systems/sap-landscapes.html

PRIMEFLEX for VMware vSAN

Fujitsu is one of the first manufacturers to receive SAP certification of a hyper-converged infrastructure (HCI) based on VMware vSAN with a 4-socket server for SAP HANA. The FUJITSU Integrated System PRIMEFLEX for VMware vSAN is based on VMware vSphere, the leading virtualization platform. The native vSAN in vSphere is capable of using all 4 CPU sockets and the 6 TB of memory in a PRIMERGY RX4770 M4 server when running SAP HANA. In other words, most customers can run their SAP HANA production databases with only one node. PRIMEFLEX for VMware vSAN thus offers the highest levels of performance and cost-effectiveness for software-defined SAP operation. For more information refer to:
https://sp.ts.fujitsu.com/dmsp/Publications/public/wp-sap-hci-ww-en.pdf

PRIMEFLEX for Nutanix Enterprise Cloud

Digital business demands are outpacing IT's ability to deliver at the speed required by the business, particularly when the data center is hardware-defined and limited to an on-premises silo. To avoid disparate IT infrastructure that increases in complexity as businesses grow, data center customers increasingly see multi-cloud, software-defined systems as the go-to-alternative.
Fujitsu's PRIMEFLEX for Nutanix Enterprise Cloud is a turnkey integrated system that includes all the hardware and software to simplify deployment across a range of choices from private, public and edge cloud infrastructure. It combines high performance, energy efficient Fujitsu PRIMERGY standard x86 servers and the software-defined server and storage technology combined in the Nutanix Acropolis cloud OS and Prism management suite to reduce complexity and TCO in a multi-cloud infrastructure environment.
If your organization is making the move to a multiple cloud deployment potentially spanning private cloud, public cloud and down to the edge, PRIMEFLEX for Nutanix Enterprise Cloud will make it a success. For more information refer to:
https://www.fujitsu.com/global/products/computing/integrated-systems/nutanix.html
https://sp.ts.fujitsu.com/dmsp/Publications/public/eb-pf4nec-sap-hana-ww-en.pdf

FUJITSU SystemInspection Service for SAP solutions

Through a series of very specific steps - from an initial analysis of your current operations to intense customer workshops and clear guidelines for your move to SAP HANA and/or related solutions - we ensure that your current workload is carefully analyzed and audited. Therefore, we can get a true picture of how your processes are performing, measure your workloads accurately, see how you consume computing power and services, and understand how it's all distributed. Then we match that data to your business strategy and requirements. For more information refer to:
https://sp.ts.fujitsu.com/dmsp/Publications/public/fl-sap-systeminspection-ww-en.pdf
https://www.fujitsu.com/emeia/products/computing/storage/systeminspection-service/

Conclusion

The purpose of Fujitsu's SAP Data Intelligence reference architecture is to provide customers with a framework supporting new operating models whilst coping with the need to bring data closer to the business. SAP Data Intelligence allows to co-ordinate and manage distributed SAP landscapes through the container-based Kubernetes platform. The underlying components, such as engines, agents and metadata storage in SAP Data Intelligence, can be executed in an isolated environment within Kubernetes thus simplifying installation, accelerating application delivery and responding flexibly to rapidly changing demands.

By delivering services and enhancements to enrich your SAP ecosystem, Fujitsu and SUSE **simplify** digital transformation by **modernizing** SAP environments and **accelerating** your success. We focus on driving value from your investments in SAP technologies, applications and systems. SUSE and Fujitsu offer SAP customers the expertise and experience to ensure top performance, reliability and security of the most critical, valuable workloads.

References

SAP Data Intelligence

SAP Blogs | what-is-data-intelligence
SAP Blogs | Introduction to SAP Data Intelligence
SAP Blogs | SAP Data Intelligence – connecting the Dots
SAP Blogs | Unified Data Integration for SAP
SAP Blogs | Debriefing: SAP Community Call "SAP Data Intelligence – Hybrid Data Management"
SAP Blogs | Want to Build a Resilient Business? Renew Your Focus on Connecting Data
SAP | SAP Data Intelligence - Machine Learning and Data Integration Services
SAP | SAP Road Map Explorer
SAP | SAP Data Intelligence - business cases
SAP | Turn Data Chaos Into Data Value
SAP Data Intelligence | SAP Community
SAP Data Intelligence | SAP | SAP Blogs
SAP Blogs | SAP Data Intelligence Trial Edition 3.1
SAP | SAP Data-Intelligence considerations for cloud or on-premises installation.
SAP | On-premise SAP Data Intelligence Memory Calculator
SAP | SAP Data Intelligence 3.14 Sizing Guide
SAP | SAP Data Intelligence - Installation Guide V. 3.14
SAP | Installation Guide for SAP Data Intelligence on-premise - Planning your installation
SAP | SAP Data Intelligence Sizing & Installation Guides,
SAP | SAP Product Availability Matrix,
SAP | Note 2693555 - SAP Data Hub 2 and SAP Data Intelligence 3: Supported Remote Systems and Data Sources
SAP | Note 2871970 - Prerequisites for installing SAP Data Intelligence 3
SAP | Note 3038897 - Installation of SAP Data Intelligence 3 on Rancher Kubernetes Engine 2
SAP | Note 2856760 - Multiple SAP Data Hub installations in one Kubernetes cluster
SAP | SAP Data Intelligence Sizing Guide Appendix B
SAP Blogs | How to install SAP Data Intelligence 3.0 on-premise edition
SAP | 2779240 - Workload-based sizing for virtualized environments
SAP | 2015392 - VMware recommendations for latency-sensitive SAP applications
SAP | 1492000 - General Support Statement for Virtual Environments
SAP | 2161991 - VMware vSphere configuration guidelines
SAP | SAP on VMware vSphere
SAP | SAP HANA Guidelines for running virtualized
SAP | Known Support Issues in Virtual Environments

SUSE/Rancher/Longhorn

Rancher | Rancher strengthens Kubernetes
SUSE | SAP Data Intelligence supported on Rancher Kubernetes Engine
SUSE Rancher | Rancher vs. RKE: What Is the Difference?
SUSE Rancher | RKE2 - How is this different from RKE or K3s?
SUSE Rancher | RKE requirements
SUSE Rancher | Recommended Cluster Architecture
SUSE Rancher | Managed Kubernetes Cluster Operations
SUSE Rancher | Hardware Requirements
SUSE Rancher | Installing Rancher in a vSphere Environment
SUSE Rancher | Best Practices for Rancher Managed vSphere Clusters
Longhorn | What is Longhorn?
SUSE Rancher | Longhorn - Persistent Volumes

NetApp

NetApp | NetApp Trident Deployment & Configuration
NetApp | NetApp Trident for Kubernetes - Persistent Volumes
NetApp | Documentation on the possible parameters of a storage class
NetApp | Examples of different use cases of Trident
NetApp | NetApp StorageGRID - Object Storage
NetApp | Introduction to StorageGrid
NetApp | Deploying StorageGRID in a Kubernetes cluster
NetApp | NetApp StorageGRID data sheet

Native Kubernetes/Docker
Docker | Docker Website
Kubernetes | Kubernetes Website
Kubernetes | Creating Highly Available clusters with kubeadm
Kubernetes| services networking ingress
etcd | etcd hardware requirements

FUJITSU PRIMERGY Rack Technology
https://www.fujitsu.com/global/products/computing/servers/primergy/racks/

FUJITSU Server PRIMERGY Rack Server
Fujitsu | Fujitsu PRIMERGY Server Models
Fujitsu | Rackserver PRIMERGY RX2530 M6
Fujitsu | Rackserver PRIMERGY RX2540 M6
Fujitsu | Multi-node Server PRIMERGY CX400 M6
Fujitsu | Multi-node Server PRIMERGY CX2560 M6

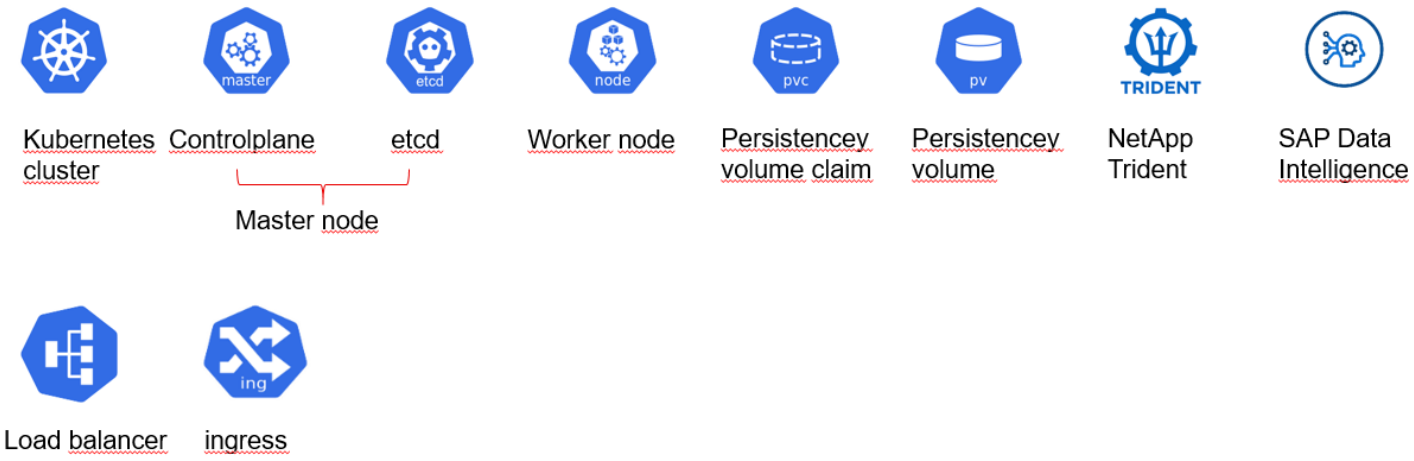
FUJITSU Server PRIMERGY Scale-out Systems
https://www.fujitsu.com/global/products/computing/servers/primergy/scale-out/
https://www.fujitsu.com/global/products/computing/servers/primergy/scale-out/cx400m4/
https://www.fujitsu.com/global/products/computing/servers/primergy/scale-out/cx2560m5/

FUJITSU Integrated System PRIMEFLEX
https://www.fujitsu.com/global/products/computing/integrated-systems/
https://www.fujitsu.com/global/products/computing/integrated-systems/sap/index.html
PRIMEFLEX for SAP HANA
https://www.fujitsu.com/global/products/computing/integrated-systems/sap-hana.html
PRIMEFLEX for SAP Landscapes
https://www.fujitsu.com/global/products/computing/integrated-systems/sap-landscapes.html
PRIMEFLEX for HPC und Big Data Analytics
https://www.fujitsu.com/global/products/computing/integrated-systems/analytics/index.html
FUJITSU IoT Solution INTELLIEDGE
INTELLIEDE Appliance A700
https://www.fujitsu.com/global/products/computing/pc/edge-computing/intelliedge-a700-appliance/
INTELLIEDGE Gateway G700
https://www.fujitsu.com/global/products/computing/pc/edge-computing/intelliedge-g700-gateway/
IP Network Switches
https://www.fujitsu.com/global/products/computing/storage/switches/ethernet-fabric/index.html

FUJITSU Services for SAP solutions
https://www.fujitsu.com/global/services/application-services/enterprise-applications/sap/

Contact
FUJITSU Technology Solutions GmbH
Address: Mies-van-der-Rohe-Str.8
80807 Munich, Germany

E-mail: cic@ts.fujitsu.com
Website: www.fujitsu.com/primeflex/sap
2021-10-06 WW-EN


master
etcd
node
pvc
pv
TRIDENT
Kubernetes cluster
Controlplane
etcd
Worker node
Persistencey volume claim
Persistencey volume
NetApp Trident
SAP Data Intelligence
Master node
ing
Load balancer
ingress