

Planning and Architecture SAP Data Intelligence on Tanzu Reference Architecture

Tanzu Solutions Engineering

September 2022

SAP Data Intelligence on Tanzu Reference Architecture

Disclaimer:

- This is a draft and contents are subject to change
- Supported version is SAP Data Intelligence 3.3
- Certification is still pending

Architecture Overview

Bill of Materials

Design Use-Cases

Core Infrastructure Requirements

<u>DNS</u>

Container Registry

S3 Compatible Object Store

<u>PKI</u>

Cluster Requirements

General Infrastructure Recommendations

Development / Test Environment

Architectural View

Production Environment

<u>Management</u>

Backup and Recovery

Container Registry

<u>Observability</u>

Logging

Architecture Overview

This document details a reference architecture for deploying SAP Data Intelligence on Tanzu Kubernetes Grid (TKG). This reference will cover topics such as Kubernetes requirements and other components needed for a successful install of SAP Data Intelligence.

This architecture should give you a path to creating a highly available, productiongrade deployment of SAP Data Intelligence. However you should not feel constrained by this exact path if your specific use cases lead you to a different deployment architecture. Design decisions in this architecture reflect the main design issues and the rationale behind a chosen solution path and if necessary can help provide rationale for any deviation.

)		Home SAP Data Intellig	ence — Mozilla Firefox		~ ^
Home SAP Data Intellige: ×	+				
← → œ O	A # or https://di.tanz	u.test/app/datahub-app-la	unchpad/	☆	⊚ ± ≡
SAP Data Intelligen	ce Launchpad 💌				0 м
Home					
Audit Log Viewer	Connection Management	Customer Data Export	Monitoring	Metadata Explorer	
	-				
23	ġ	e +	R		
System Management	ML Scenario Manager	Modeler			
dī.	<u>k</u>	ne -			

Figure 1: SAP Data Intelligence Launchpad

Bill of Materials

Component	Version
-----------	---------

Tanzu Kubernetes Grid	1.5.4
Kubernetes	1.21, 1.22
SAP Data Intelligence	3.3
VMware NSX-T Advanced Load Balancer	20.1.8
VMware vSphere/vSAN	7.0u3
VMware NSX-T	3.1.3
Harbor	2.4.2
S3 Data Store	Minio

This reference architecture was validated using these versions of Tanzu and SAP Data Intelligence solutions.

Design Use-Cases

Design #	Decision	Rationalization	Ramifications
TKG-001	А	Appropriate choice for a	Not an option for

development/test cluster	development/test, non- critical environment where availability, or loss of data is of little to no concern. Primary benefit of this design is ease of	workloads requiring high availability, durability, and resiliency.
	deployment.	

a level of availability.		TKG-002	A production cluster	Development/test or production workloads that need to scale, and require a level of availability.	
--------------------------	--	---------	-------------------------	--	--

Core Infrastructure Requirements

DNS

You must have a working DNS setup. SAP DI will not connect to a private container registry via an IP address alone.

SAP Data Intelligence Installation Prerequisites: Container Registry	https://help.sap.com/docs/SAP_DATA_IN TELLIGENCE_ON- PREMISE/a8d90a56d61a49718ebcb5f650 14bbe7/946d67f312e74a13942f23e50aa0 6867.html

Container Registry

A container registry needs to be provided for installation and for building models. This RA provides for the use of Harbor, an OSS container registry supported by VMware and integrated in the Tanzu product suite. If other registry providers are desired, you can follow the SAP Installation Guide for more details.

Requirements for Installing SAP Data Intelligence on Kubernetes - Container Registry	https://help.sap.com/docs/SAP_DATA_IN TELLIGENCE_ON- PREMISE/a8d90a56d61a49718ebcb5f650 14bbe7/946d67f312e74a13942f23e50aa0 6867.html

S3 Compatible Object Store

The SAP DI application uses an S3 datastore for two purposes. One is to be a backup target for the application and the second is to act as a file storage location for the SAP Data Lake (SDL) connection for any files (e.g CSV) that will be ingested into ML pipelines.

The S3 Datastore can be the same or split up per the preferences of the Application and Infrastructure Ops teams.

If you host a private S3 datastore, there are many options to choose from. Consult the VMware Marketplace for an option suitable for your implementation.

https://marketplace.cloud.vmware.com/ services?search=s3

PKI

Formal PKI management is recommended but not required for successful implementation if SAP DI.

Certificates must be valid or in the case of self-signed certificates for dev/test, must be included in the trust store by the OS, SAP DI application, and registry TLS access.

Cluster Requirements

General Infrastructure Recommendations

Storage

- vSAN
- S3 Datastore used for Application backups and file storage for SAP Data Lake.
- Container registry used for Installation and Modeling container deployment

Compute

Use the following CPU/RAM/DISK ratio as a general guideline.

• 8 vCPU / 16 GB RAM / 400 GB storage

For more information, refer to the <u>Sizing Guide for SAP Data Intelligence</u>.

Network

For ingress to the SAP DI app that is hosted in ClusterIP mode, use the <u>NSX-TALB</u> (L4) + Contour (L7) setup.

Development / Test Environment

The following components and their respective specifications are required for hosting a minimal installation of SAP Data Intelligence on Tanzu.

SAP DI is hosted on a single Kubernetes cluster, in its own namespace. It is recommended to only install a single instance of SAP DI on any one cluster and to create separate clusters for separation of dev/test and production.

1 TKG Management Cluster

- 1 control plane node
- 1 worker node

1 Tanzu Kubernetes Cluster

- 1 control plane node
- 3* worker nodes (1 node = 8 CPU, 32GB RAM, 80GB Storage)

Ingress

NSX-T Advanced Load Balancer (L4) + Contour (L7)

• Multiple envoy pods

SAP Data Intelligence

• SAP Data Intelligence 3.3

Kubernetes requirements:

- NSX Advanced Load Balancer to SAP DI
- Default storage class, default TKG 80GB is good starting point
- At least 16GB per node

Architectural View



Figure 2: Tanzu Kubernetes Grid on vSphere

Production Environment

The following components and their respective specifications are required for hosting a moderately available installation of SAP Data Intelligence.

It is generally recommended for production environments to install SAP Data Intelligence on its own dedicated cluster, which will be followed here.

SAP DI must be installed into a single Kubernetes cluster.

Tanzu Management Cluster

• 3 control plane nodes

• 3 worker nodes

Tanzu Kubernetes Cluster

- 3 control plane nodes
- 3* worker nodes (1 node = 8 CPU, 32GB RAM, 80GB Storage)

*The size of the nodes will determine the number of nodes needed, or vice versa. The larger the node size, the number of nodes can be reduced.

Ingress

NSX-T Advanced Load Balancer (L4) + Contour (L7)

• Multiple envoy pods

SAP Data Intelligence

• SAP Data Intelligence 3.3

Kubernetes requirements:

- Load Balancer to SAP Data Intelligence
- Default storage class (80GB is good starting point)
- At least 8GB per node

Management

Tanzu Mission Control (TMC) can provide TKG Management and Workload cluster lifecycle management.

TMC also has integrations with Tanzu Observability for monitoring of clusters and their workloads.

vmw Tanzu Mission Control		௴ ⊙ 🖩
ري ف Leunchood	← & sap-di-1 ● Investivy Overview Nodes Namespaces Workloads Add-one Continuous Delivery Secrets Inspections Data protect	C Last updated treinule ago
4): Outlier groups	Challer gringe 190-61 Provider vicehrer Munagemet chalter attriched Type Attriched Provideorer Jackheid Kubernetes version vi 22.9-vrimeire 1 Labelle (Int: Coud immers antri/metric traterage)	Node court 6 Created Thursday, July 14, 2022, 10 Algon Total memory INI 76 GB Total cores 36 GPUs
(ii) Namespaces	Requested/Allocatable CPU 52%	Requested/Allocatable memory 20% an era/ron via
E Workloads	Component health © controler-manager © etcl-0 © kube-spiserver ● scheduler	Worker nodes () © 3 nodes teathy
Q, Policies > () Inspections () Events	Agent and extensions health a gent optimer of health between operator between manyer between manyer between between between operator between manyer between be	Integrations ADD INTEGRATION > Name Two Adustry Integration Workley
Je Administration		① No viewable integrations.
度 Audit logs		Integrations that you have permission to stew will show up here.

Figure 3: Tanzu Mission Control

Backup and Recovery

KUBERNETES

The recommended backup solution for SAP Data Intelligence is the built-in backup to an S3 Datastore.

Container Registry

SAP DI requires a private container registry to clone images from the SAP Registry during installation to be applied to the TKG Workload clusters. In addition, the Modeling components of SAP DI will use Kaniko on the Kubernetes cluster to create containers containing models and push them to a private registry that can be the same registry as the installation registry or a different one.

<u>Harbor</u> is the recommended container registry to use with Tanzu and is supported by VMware and integrated into the Tanzu product suite. Harbor can be installed standalone in a VM or be hosted on a TKG Workload Shared Services Cluster.

Note for this validation only the standalone Harbor cluster has been tested.

↔ ⊖ C	OA	https://registry.tanzu.test/harbor/projects/13/repositories			☆	⊚ ± ≣
Harbor	QSearch	Harbor -			🛞 English 👻	& admin
🖧 Projects				Access Level	Storage use	
🗏 Logs	윩	sap-di _{System Admin}		Delivere	31.75GiB	of unlimited
& Administration				Private	01.70012	or uninnited
		nmary Repositories Helm Charts Members Labels	Scanner P2P Prehe	at Policy Robot Accou		
C Robot Accounts		analy repetitions from coming the second				
D Registries				PUSH	COMMAND - C	
Registries Cr Replications						Q ⊞≡ C
			t y Antifacts	PUSH	COMMAND - C	
Ct Replications] Name	© ₩ Årtifects T			ime
Cr Replications		Name sap-tl/com sap.txts.docker/storagegateway	1 y Arittaita 1	Puth	Last Modified Y	ime AM
CI Replications		Name sap-B/com sap.bds.docker/storagegateway sap-bl/com sap.datahub imuxk86_64/app-base	t y Artifactu 1 1 1	Puls 37	Last Modified 7 7/17/22, 8:28	ime AM AM
 G: Replications 		Name sap-dl/com sap.dxts.docker/storagegateway sap-dl/com sap.dxtahub.imuox86_64/app-base sap-dl/com sap.dxtahub.imuox86_64/app-date	T Artifacts 1 1 1 1 1 1	Puls 37 32	Last Modified T 7/17/22, 8:28 7/17/22, 8:28	ame AM AM AM
Replications Distributions Labels Project Quotas O Interrogation Services		Name sap-tll/com sap.bds.docker/storagegateway sap-tll/com sap.dxtahub.linuxx86_64/app-base sap-tll/com sap.datahub.linuxx86_64/app-data sap-tll/com sap.datahub.linuxx86_64/app-clata	 W Aritfacts 1 1 1 1 1 1 	Puts 37 22 28	Last Modified 7 7/17/22, 8:28 7/17/22, 8:28 7/17/22, 8:28	ime AM AM AM AM

Figure 5: Harbor Container Registry

Observability

The following are options for monitoring real-time metrics and statistics.

- **SAP Data Intelligence**, includes an installation of Prometheus and Grafana configured for the application.
- **Tanzu Observability**, an advanced service for collecting metrics, traces, and logs. Includes 200+ integrations with the most popular services and libraries, including Prometheus.
- **Prometheus**, component for collecting host metrics and outputting to charting components.

Logging

The SAP Data Intelligence application has a built-in Elasticsearch and Kibana instance configured to receive logs from the running pods and jobs.

In addition to the built in application capability, you may install the Tanzu packages for Fluent Bit on Tanzu Workload clusters and send logging information to an instance of vRealize Log Insight or any other endpoints that are supported.