

Document Version: 1.1 – 2022-07-06

Getting Started with SAP Data Intelligence 3.2 Trial Edition

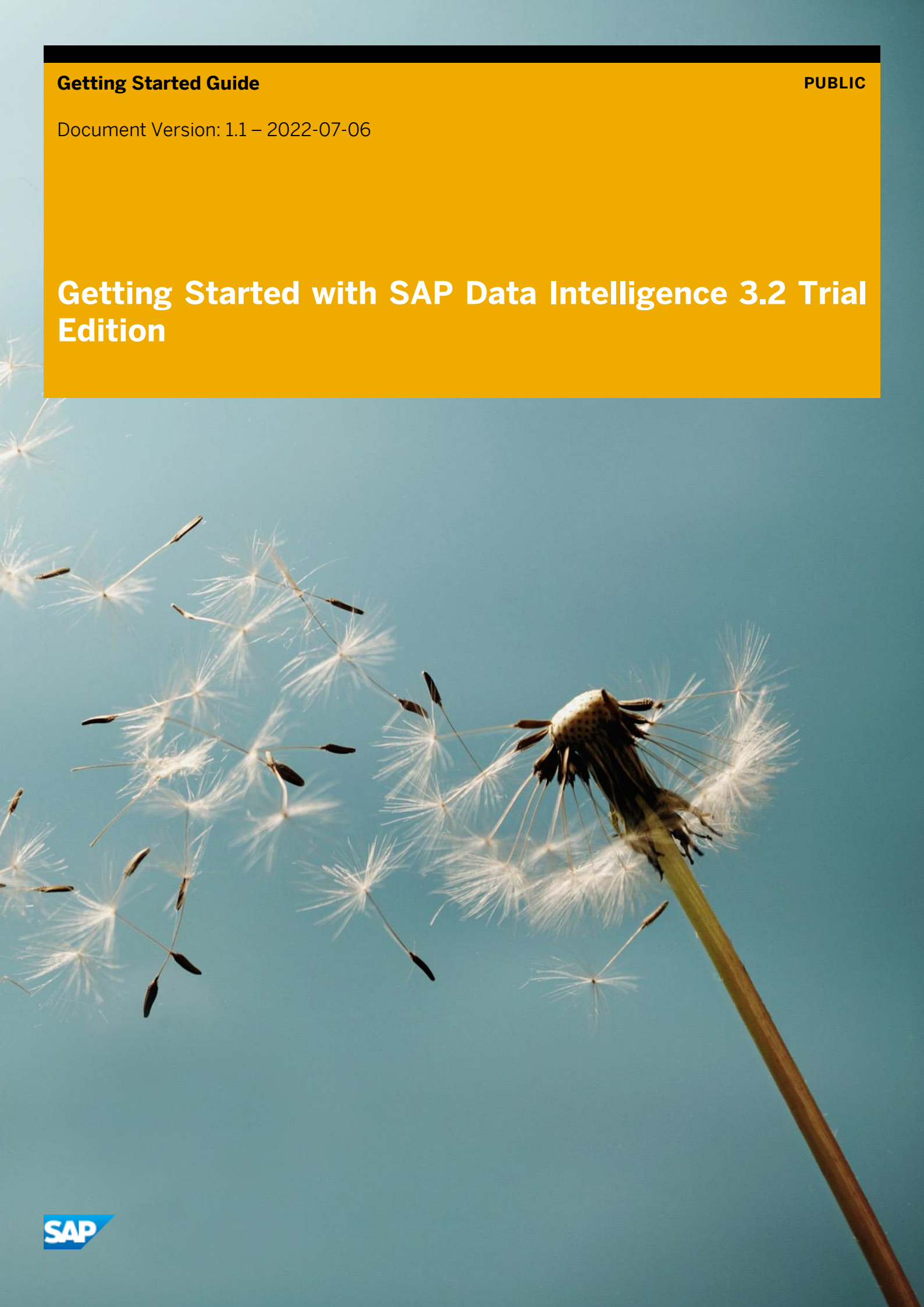


Table of Contents

1	Solution Information.....	3
1.1	Product Overview.....	3
1.2	Architecture Overview	4
1.3	Installed Product	4
2	Provision the Solution	5
2.1	Prerequisite: Cloud Account Provider.....	5
2.1.1	Google Cloud Platform	5
2.2	Accessing the SAP Cloud Appliance Library.....	6
2.3	Creating a Solution Instance	6
3	Accessing the Solution	10
3.1	Overview	10
3.2	Determining the external IP addresses of your solution instance	10
3.3	Accessing SAP Data Intelligence on Google Cloud Platform	11
3.4	Accessing the Jump Box for troubleshooting.....	12
4	Licenses	14
4.1	Running Solution with a Product License Key.....	14
4.2	Running Solution as a Trial for a Free Period.....	14
5	Security Aspects	15
6	Troubleshooting	16
6.1.1	Pipeline graphs.....	16
6.1.2	Obtaining kubeconfig file.....	16
6.1.3	Accessing SAP Data Intelligence namespace	16
6.1.4	Accessing application logs.....	17
6.1.5	Enable access to Kubernetes Dashboard.....	18
7	Appendix.....	19

1 Solution Information

1.1 Product Overview

SAP Data Intelligence, trial edition is a pre-configured appliance for evaluating and testing SAP Data Intelligence, the successor to SAP Data Hub.

In addition to all the data sharing, pipelining and orchestration capabilities that were originally featured in SAP Data Hub, SAP Data Intelligence introduces a way to manage machine learning models with built-in tools for data governance, management, and transparency:

- Data Pipelines are data-driven applications consisting of reusable and configurable operators.
- Data Workflows orchestrate processes across a data landscape.
- Data Governance allows you to extract metadata from connected data stores. It supports discovery, profiling and search.
- Automated Machine Learning (AutoML) lets you automate the creation of complex machine learning pipelines with intuitive tools for configuring machine learning artifacts.

To learn more about SAP Data Intelligence, refer to:

- [What's new in SAP Data Intelligence 3.2](#)
- [SAP Data Intelligence Walkthrough](#)
- [Product Home Page](#)
- [Official documentation](#)
- Hands-on learning material:
 - [SAP Developer Center](#) (Create and Manage End-to-End Data Pipelines with SAP Data Intelligence)
 - TechEd Workshop: [DAT163 - Catalog, Discover, and Prepare: Data Governance with SAP Data Intelligence](#)

1.2 Architecture Overview

The SAP Data Intelligence, trial edition currently is available on Google Cloud Platform. While creating a solution instance of SAP Data Intelligence, trial edition, the SAP Cloud Appliance library creates:

1. Jump Box to orchestrate the instances creation
2. Kubernetes cluster to run SAP Data Intelligence

The Jump Box is used during the creation of the solution instance to run the installer for SAP Data Intelligence. As a user of the SAP Data Intelligence, trial edition you typically do not need to access the Jump Box except for troubleshooting (in rare cases). The Jump Box is automatically terminated 4 hours after the installation has finished successfully (afterwards it cannot be accessed anymore).

SAP Data Intelligence leverages container technology, in particular Docker and Kubernetes, that is offered by the cloud hyperscaler vendor. It allows to query large amounts of data (in distributed storages like Google Cloud Storage (GCS)) and to run highly scalable data flows and data-driven applications. As part of this pre-configured appliance, SAP Data Intelligence uses a Kubernetes cluster with three worker nodes.

1.3 Installed Product

The SAP Data Intelligence, trial edition leverages release 3.2, Full platform stack

2 Provision the Solution

2.1 Prerequisite: Cloud Account Provider

To use the SAP Cloud Appliance Library in order to create a solution instance of SAP Data Intelligence, trial edition, you need to have access to a Google Cloud Platform project. You need to link your account in the SAP Cloud Appliance Library to those.

2.1.1 Google Cloud Platform

If you like to use Google Cloud Platform, you need to enable certain APIs and to create a service account with certain roles.

Required APIs

To create a solution instance of SAP Data Intelligence, you need to enable the following APIs in your Google Cloud Platform project.

Name	Description
Cloud Resource Manager API	The Google Cloud Resource Manager API provides methods for creating, reading, and updating project metadata.
Compute Engine API	Creates and runs virtual machines on Google Cloud Platform.
Kubernetes Engine API	The Google Kubernetes Engine API is used for building and managing container-based applications, powered by the open source Kubernetes technology.
Container Registry API	Google Container Registry provides secure, private Docker image storage on Google Cloud Platform. Our API follows the Docker Registry API specification, so we are fully compatible with the Docker CLI client, as well as standard tooling using the Docker Registry API.
Google Cloud Storage	Google Cloud Storage is a RESTful service for storing and accessing your data on Google's infrastructure.

Service account and roles

The service account which you use to access your Google Cloud Platform project through the SAP Cloud Appliance Library needs to have the following roles.

Name	Description
Compute Instance Admin (v1)	Full control of Compute Engine instances, instance groups, disks, snapshots, and images. Read access to all Compute Engine networking

Name	Description
	resources.
Compute Network Admin	Full control of Compute Engine networking resources.
Compute Security Admin	Full control of Compute Engine security resources.
Kubernetes Engine Admin	Full management of Kubernetes Clusters and their Kubernetes API Objects.
Service Account User	Create VMs and other GCP tasks with a service account. Users cannot impersonate the account directly as they can with Service Account Actor role.
Storage Admin	Full control of GCS resources.

For more information about how to link the SAP Cloud Appliance Library with a Google Cloud Platform project, refer to the [FAQ page](#).


2.2 Accessing the SAP Cloud Appliance Library

Open the SAP Cloud Appliance Library in your web browser using the following link: <https://cal.sap.com>

If you are a first-time user of SAP Cloud Appliance Library, familiarize yourself with its basic concepts and how to work with the user interface by reading the [documentation](#).

2.3 Creating a Solution Instance

The following steps guide you through the creation of a solution instance of SAP Data Intelligence, trial edition. The screenshots have been taken for Google Cloud Platform. Wherever you need more information, refer to the [documentation](#).

1. Navigate to solution SAP Data Intelligence, trial edition 3.2
2.  Press the Create Instance button and then switch to Advanced Mode (pressing the button in the lower right of the screen).
3. Select the cloud provider account. Move to step 2 by pressing the Step 2 button.
4. Enter Name, Number of Instances (1), Region, Zone¹, Network and Subnet. Ensure that the checkbox Public

Static IP Address is marked. Move to step 3 by pressing the Step 3 button.

2. Instance Details

Enter the general properties of the solution instance:

*Name:

Description:

*Number of Instances: ⓘ

*Region: ▾

*Zone: ▾

*Network: ▾

*Subnet: ▾

Public Static IP Address ⓘ

Step 3

5. Next check the parameters of the size of the Jump Box. You can leave this at default unless you are having problems with the deployment. Move to step 4 by pressing the Step 4 button.

¹ only on Google Cloud Platform

6. To restrict public access to the solution it is highly recommended to limit the IP range of the HTTPS access point to your own IP address or subnet. To identify your IP address, go to <https://ifconfig.co/> and add /0 to the end of it, e.g. 12.34.56.78/0. This setting can also be changed after the solution has been deployed.

< Advanced Mode: Create Instance

4. Kubernetes Cluster

Number of Nodes:

Size: ▾

Access Points

Service	Port Range	IP Range	Enabled
<input type="text" value="HTTPS"/> ▾	<input type="text" value="443"/>	<input type="text" value="<your-ip-address>/0"/>	<input checked="" type="checkbox"/>

Type: Default

Step 5

7. Move to step 5 by pressing the Step 5 button.

- Next define a solution password. We will refer to this as <Master Password> for the rest of this document. Move to step 6 by pressing the Step 6 button.

1 Account Details — 2 Instance Details — 3 Virtual Machines — 4 Kubernetes Cluster — 5 Solution Password

5. Solution Password

Set the master password for the solution instance. You can check the Getting Started Guide to learn where to use this password.

*Password:

*Retype Password:

i

The valid characters are: A-Z, a-z, 0-9, \$, #, _

The first character has to be one of the following: A-Z, a-z, \$, #

The first 3 characters cannot be one and the same

The password must contain at least 1 lowercase letter(s)

The password must contain at least 1 uppercase letter(s)

The password must contain at least 1 digit(s)

The password must be between 8 and 9 characters long

Step 6

- Finally set up the schedule of the solution instance. You can define a schedule to suspend it. And you can set a termination date. Finally press the Review button.

- Assuming you are happy with all parameters press the Create button. It will take approximately 2 hours to create the solution instance. The startup time strongly depends on cloud provider and region.

You can store and/or download the private key. This is needed in case you want to access the Jump Box on operating system level which is typically only needed for troubleshooting activation issues. It cannot be used to access the Kubernetes cluster nodes.

6. Schedule Details

Set up the schedule of the solution instance, when to suspend it, and when to terminate it.

Time Zone: (GMT+02:00) Amsterdam, Berlin, Bern, Rome, St... 

Scheduling Options

- Activate and suspend by schedule
- Suspend on an exact date
- Manually Activate and Suspend

Suspend Date: Sep 28, 2018  21:21:32  

Termination Date

- Set the termination date of the solution instance

[Review](#)

3 Accessing the Solution

3.1 Overview

As already described in chapter 1

[Architecture Overview](#), a solution instance of SAP Data Intelligence, trial edition consists of:

- A temporary virtual machine to orchestrate the solution instance creation (Jump Box)
- A Kubernetes cluster to run SAP Data Intelligence

Subsequently you learn how to access your solution instance.

- Chapter 3.2 tells you how to determine the external IP addresses of your solution instance.
- Chapter 3.3 shows how you can access SAP Data Intelligence on Google Cloud Platform.
- Chapter 6 outlines how to access the Jump Box for troubleshooting.

3.2 Determining the external IP addresses of your solution instance

If you like to access your solution instance, you might need to know the external IP addresses or domain names. You can look them up by opening your solution instance in the SAP Cloud Appliance Library:

The following IP addresses / domain names are available in detail:

	Google Cloud Platform
SAP Data Intelligence External IP Address	X
Kubernetes Master IP Address	X
Jump Box Internal IP Address	X
Jump Box External IP Address	X
Transaction Coordinator Internal IP Address	X
HANA Internal IP Address	X

Instances / My SAP Data Intelligence Trial Download Key Connect Suspend Edit ? ...

Owned By: Georgi Stankov (1342828) ■ Active Cost Forecast ⓘ
Disclaimer
1.04 USD per hour when Active
52.43 USD per month when Suspended

Created On: Jan 8, 2021, 16:18:08 GMT+02:00
Last Activation At: Jan 8, 2021, 16:24:26 GMT+02:00
Next Suspend At: Jan 9, 2021, 00:17:38 GMT+02:00

INFO SOLUTION INFO VIRTUAL MACHINES KUBERNETES CLUSTER SAP DATA INTELLIGENCE 3.0 SCHEDULE

Account: Antonia GCP	Region: europe-west1	<input type="checkbox"/> Public Static IP Address	Endpoints 34.120.49.143 SAP Data Intelligence External IP Address 34.78.177.0 Kubernetes Master IP Address 10.132.0.29 Jump Box Internal IP Address 35.205.27.137 Jump Box External IP Address
Cloud Provider: Google Cloud Platform	Zone: Zone B	<input type="checkbox"/> Termination Protection	
	Network: ECS Standard-VPC		
	Subnet: ecs-standard-vpc		

3.3 Accessing SAP Data Intelligence on Google Cloud Platform

On Google Cloud Platform the SAP Data Intelligence, trial edition application is not assigned a hostname and must be accessed through the web browser by using an IP address. computer:

Instances / My SAP Data Intelligence Trial Download Key Connect Suspend Edit ? ...

Owned By: Georgi Stankov (1342828) ■ Active Cost Forecast ⓘ
Disclaimer
1.04 USD per hour when Active
52.43 USD per month when Suspended

Created On: Jan 8, 2021, 16:18:08 GMT+02:00
Last Activation At: Jan 8, 2021, 16:24:26 GMT+02:00
Next Suspend At: Jan 9, 2021, 00:17:38 GMT+02:00

INFO SOLUTION INFO VIRTUAL MACHINES KUBERNETES CLUSTER SAP DATA INTELLIGENCE 3.0 SCHEDULE

1. Press the Connect button in the SAP Cloud Appliance Library
2. In the popup window press Open

Connect to the Instance

Select an access point to connect to the solution.

Service	Target	Operation
HTTPS	SAP Data Intelligence	Open

Close

- Your (default) web browser opens. The Data Intelligence application is signed using a self-signed certificate and for that reason you will see an error message that the web browser is not able to verify the security certificate. Confirm this error message and choose to trust this certificate. Logon using the credentials listed below.

Name	Value	Description
URL	https://<IP-address>	SAP Data Intelligence logon screen
Tenant name	default system	All standard applications are in the <code>default</code> tenant. The <code>system</code> tenant is only used for administration tasks.
Username	system	SAP Data Intelligence user for default tenant SAP Data Intelligence user for the system tenant
Password	<Master Password>	The initial master password of the system you provided in the SAP Cloud Appliance Library when creating the instance.

3.4 Accessing the Jump Box for troubleshooting

You can access the Jump Box on operating system level via the secure shell protocol. This is only necessary for troubleshooting. For a detailed description, go to [this FAQ wiki page](#) and check this question: How to connect to a running instance via the secure shell protocol (SSH)?

On operating system level the following users and user groups are available:

Name	Description
root	super user with access to all commands and files

Attention: The Jump Box is automatically terminated 4 hours after the installation has finished successfully (afterwards it cannot be accessed anymore).

Checking the logs of the activation process

The progress of the activation can be tracked in the `/var/log/applianceagent.log` file. You can stream the current activity by using below command:

```
tail -f /var/log/applianceagent.log
```

Using kubectl on the Jump Box

The Kubernetes command line interface (kubectl) is preinstalled on the Jump Box. This allows you to easily connect to the containers (pods) of the Kubernetes cluster used by SAP Data Intelligence.

If you, for example, run a `kubectl get nodes` command, you get a list of the nodes (virtual machines) of the

Kubernetes cluster.

```
lkgint-166738-@ -jump-box:~ # kubectl get nodes
NAME                                STATUS    ROLES    AGE     VERSION
gke-lkgint-166739-@ -clus-sapcal-d7ae12e2-21m1  Ready    <none>   2h     v1.8.9-gke.1
gke-lkgint-166739-@ -clus-sapcal-d7ae12e2-pzj4  Ready    <none>   2h     v1.8.9-gke.1
gke-lkgint-166739-@ -clus-sapcal-d7ae12e2-vg6f  Ready    <none>   2h     v1.8.9-gke.1
lkgint-166738-@ -jump-box:~ #
```

For more information refer to the [Kubernetes documentation](#).

4 Licenses

4.1 Running Solution with a Product License Key

To use the instance created from this solution under your own SAP Product License Agreements, you have to unlock the solution in SAP Cloud Appliance Library. As a prerequisite, you must own the required SAP Product licenses and need to purchase the SAP Cloud Appliance Library subscription package. For more information, see [Unlocking Solutions](#).

4.2 Running Solution as a Trial for a Free Period

This solution can be used with a Free Trial License Agreement for the configured free period. You will be asked to agree to the Free Trial License Agreement during the instantiation steps.

Once you create the solution instance, the SAP system will generate a 30-day temporary license key that is sufficient for exploration purpose. After the expiration of the pre-installed temporary license a valid license is required to keep using the solution instance.

Note that you are not allowed to install license keys under the Free Trial License Agreement. You can continue the solution with the Product License Agreement any time prior to the end of the free period by unlocking it.

5 Security Aspects

Be aware that creating your instances in the public zone of your cloud computing platform is convenient but less secure. Ensure that only port 22 (SSH) is opened when working with Linux-based solutions and port 3389 (RDP) when working with Windows based solutions. In addition, we also recommend that you limit the access to your instances by defining a specific IP range in the Access Points settings, using [CIDR notation](#). The more complex but secure alternative is to set up a virtual private cloud (VPC) with VPN access, which is described in [this tutorial on SCN](#).

The list below describes the ports opened for the security group formed by the server components of your solution instance:

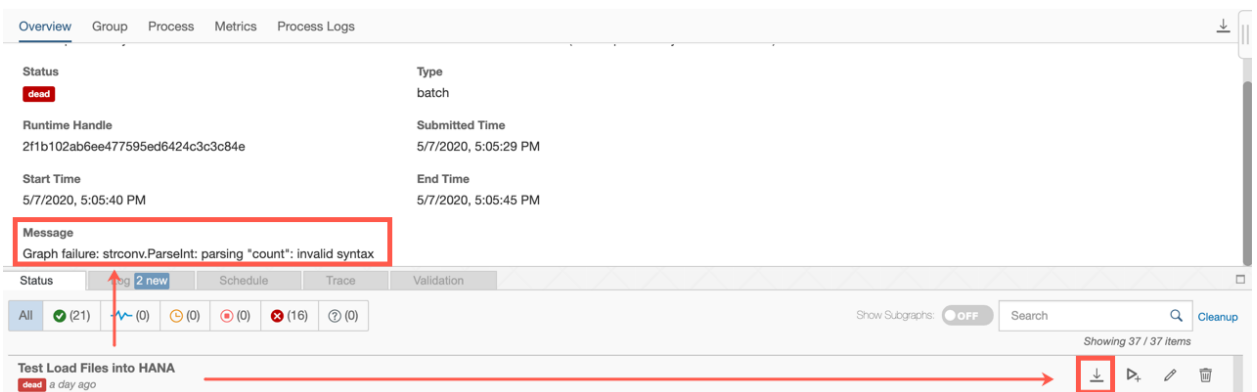
Protocol	Port	Description
SSH	22	Used for SSH connection to Jump Box
RDP	3389	Used for remote desktop connection to Windows Remote Desktop
TCP	443	Used for HTTP(S) connection to SAP Data Intelligence

For more information about security vulnerabilities, see [this document](#).

6 Troubleshooting

6.1.1 Pipeline graphs

In the Pipeline Modeler application you may view the error message of all dead graphs by clicking on the title of the graph which open an overview tab. Alternatively you may download a Diagnostic Information of the dead graph as a .tar.gz archive file by clicking on the download button to the far right of the graph title.



6.1.2 Obtaining kubeconfig file

To collect logs from the underlying Kubernetes pods/containers you will first need a workstation or jumpbox with the [kubectl](#) command line tool installed.

Each cloud vendor has different mechanisms to download the kubeconfig file for your Kubernetes cluster that is running the SAP Data Intelligence Trial Edition.

For Google Cloud: If you deployed the Trial Edition in Google Cloud then you have the [option of using a preconfigured jumpbox](#). If you wish to use your own workstation then you will have to install [Google Cloud SDK](#) or access to [Google Cloud Shell](#), then run the command [gcloud container clusters-get credentials](#)

6.1.3 Accessing SAP Data Intelligence namespace

Regardless of cloud vendor, the SAP Data Intelligence will be deployed in the Kubernetes namespace with the name `datahub`


```
kubectl get namespaces
```

NAME	STATUS	AGE
datahub	Active	23d
datahub-system	Active	22d
default	Active	23d
ingress-nginx	Active	22d
kube-node-lease	Active	23d
kube-public	Active	23d
kube-system	Active	23d

6.1.4 Accessing application logs

The following table illustrate most common used pod names and how they relate to the individual applications in SAP Data Intelligence:

Application name	Pod name (Excluding suffix identifier)
Pipeline Modeler	pipeline-modeler (Front end) vflow-graph (Pipeline execution instance)
Metadata Explorer	datahub-app-data
Connection Manager and backend connectivity components	[Front end] datahub-app-core [HANA / S3 / GCS] Flowagent [ABAP only] axino-service [ADL / HDSF] storagegateway
Vora Disk engine	vora-disk-0
Issues with building docker images	See Kaniko pod logs and Pipeline Modeler logs

Note that some applications are user specific, while others are shared across the tenant. For that reason, you may see multiple pods that belong to the same application.

To view a complete list of pods deployed in your Kubernetes cluster use the command:

```
kubectl get pods --namespace datahub
```

To view the logs of a pod use the command:

```
kubectl logs <pod-name-with-suffix> --namespace datahub
```

6.1.5 Enable access to Kubernetes Dashboard

If you prefer interacting with the Kubernetes cluster using a web user interface instead of using the `kubectl` command line tool then you may consider deploying the Kubernetes Dashboard. A detailed step by step guide can be found on the official [SAP Data Intelligence Administration guide](#). Note that the Kubernetes Dashboard is an open-source tool and not created or maintained by SAP.

Caution: Exposing the Kubernetes Dashboard to the public internet poses significant security risks to your Kubernetes cluster.

7 Appendix

Installed Software Components

Name	Release	Support Package Stack	Patch
SAP DATA INTELLIGENCE Platform full platform stack	3	2	7

Detailed information (incl. licenses) about 3rd party software used can be found in the [Free & Open Source Notices](#).



www.sap.com/contactsap

© 2022 SAP SE or an SAP affiliate company. All rights reserved.
No part of this publication may be reproduced or transmitted in any form or for any purpose without the express permission of SAP SE or an SAP affiliate company.

The information contained herein may be changed without prior notice. Some software products marketed by SAP SE and its distributors contain proprietary software components of other software vendors. National product specifications may vary.

These materials are provided by SAP SE or an SAP affiliate company for informational purposes only, without representation or warranty of any kind, and SAP or its affiliated companies shall not be liable for errors or omissions with respect to the materials. The only warranties for SAP or SAP affiliate company products and services are those that are set forth in the express warranty statements accompanying such products and services, if any. Nothing herein should be construed as constituting an additional warranty.

SAP and other SAP products and services mentioned herein as well as their respective logos are trademarks or registered trademarks of SAP SE (or an SAP affiliate company) in Germany and other countries. All other product and service names mentioned are the trademarks of their respective companies. Please see www.sap.com/corporate-en/legal/copyright/index.epx for additional trademark information and notices.