Getting Started Guide **CUSTOMER**

Document Version: 1.0 – 2018-06-29

Getting Started with SAP Data Hub 1.0 SPS 04, trial edition

Table of Contents

1 Solution Information

1.1 Product Overview

The SAP Data Hub, trial edition is a pre-configured appliance for evaluating and testing SAP Data Hub.

SAP Data Hub includes data sharing, pipelining and orchestration capabilities that help companies accelerate and expand data flow across a diverse data landscape:

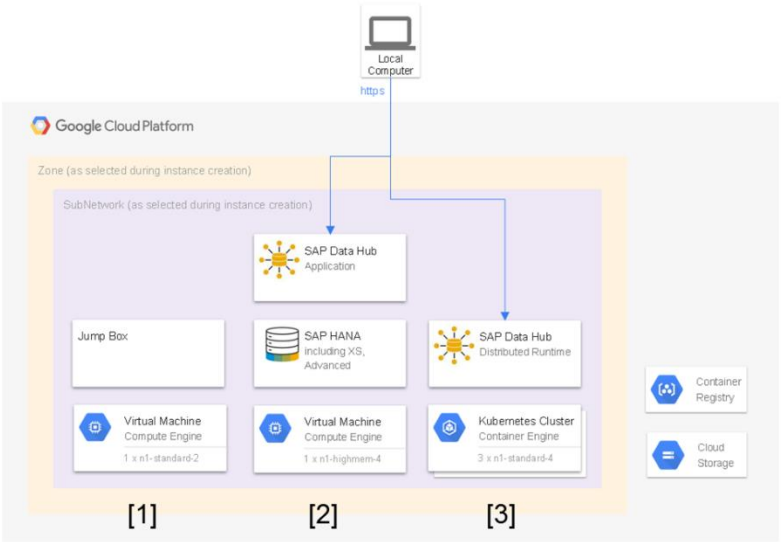
- Data Pipelines are data-driven applications consisting of reusable and configurable operators.
- Task Workflows orchestrate processes across a data landscape.
- And Data Governance allows you to extract metadata from connected data stores. It supports discovery, data quality/integration, profiling and search.

To learn more about SAP Data Hub, refer to:

- Product Home Page: https://www.sap.com/products/data-hub.html
- SAP Help Portal: https://help.sap.com/viewer/p/SAP_DATA_HUB
- Tutorials: https://www.sap.com/developer/topics/data-hub.html
- Frequently Asked Questions: https://blogs.saphana.com/2017/10/04/what-is-sap-data-hub-and-answers-to-other-frequently-asked-questions/

1.2 Architecture Overview

While creating a solution instance of SAP Data Hub, trial edition, the SAP Cloud Appliance library creates:



- a virtual machine to orchestrate the solution instance creation (Jump Box) [1]
- a virtual machine to run the SAP Data Hub Application [2]
- a Kubernetes cluster to run the SAP Data Hub Distributed Runtime [3]

The Jump Box [1] is a small virtual machine (2 cores, 7.50 GB RAM). It is used during the creation of the solution instance to run the installer for the SAP Data Hub Distributed Runtime. As a user of the SAP Data Hub, trial edition you typically do not have to access the Jump Box except for troubleshooting (in rare cases).

The SAP Data Hub Application [2] is built on top of SAP HANA and SAP HANA Extended Application Services, advanced model. It is a rather lightweight application that serves as the central entry point for end-users of SAP Data Hub. It provides you with a single point of access to a range of tools. As part of this pre-configured appliance the SAP Data Hub Application runs on a virtual machine with 4 cores and 26GB memory.

The SAP Data Hub Distributed Runtime [3] leverages container technology, in particular Docker and Kubernetes. It allows to query large amounts of data (in distributed storages like Amazon S3, Azure Data Lake (ADL), Azure Storage Blobs (WASB), HDFS, Google Cloud Storage) and to run highly scalable data flows and data-driven applications. As part of this pre-configured appliance the SAP Data Hub Distributed Runtime uses a Kubernetes cluster with three worker nodes (each 4 cores, 15 GB RAM).

1.3 Installed Products

The SAP Data Hub, trial edition consists of:

- SAP Data Hub 1.0 SPS 04
    - SAP Data Hub Application
    - SAP Data Hub Distributed Runtime
- SAP HANA, express edition 2.0

2 Provisioning the Solution

2.1 Prerequisite: Cloud Provider Account

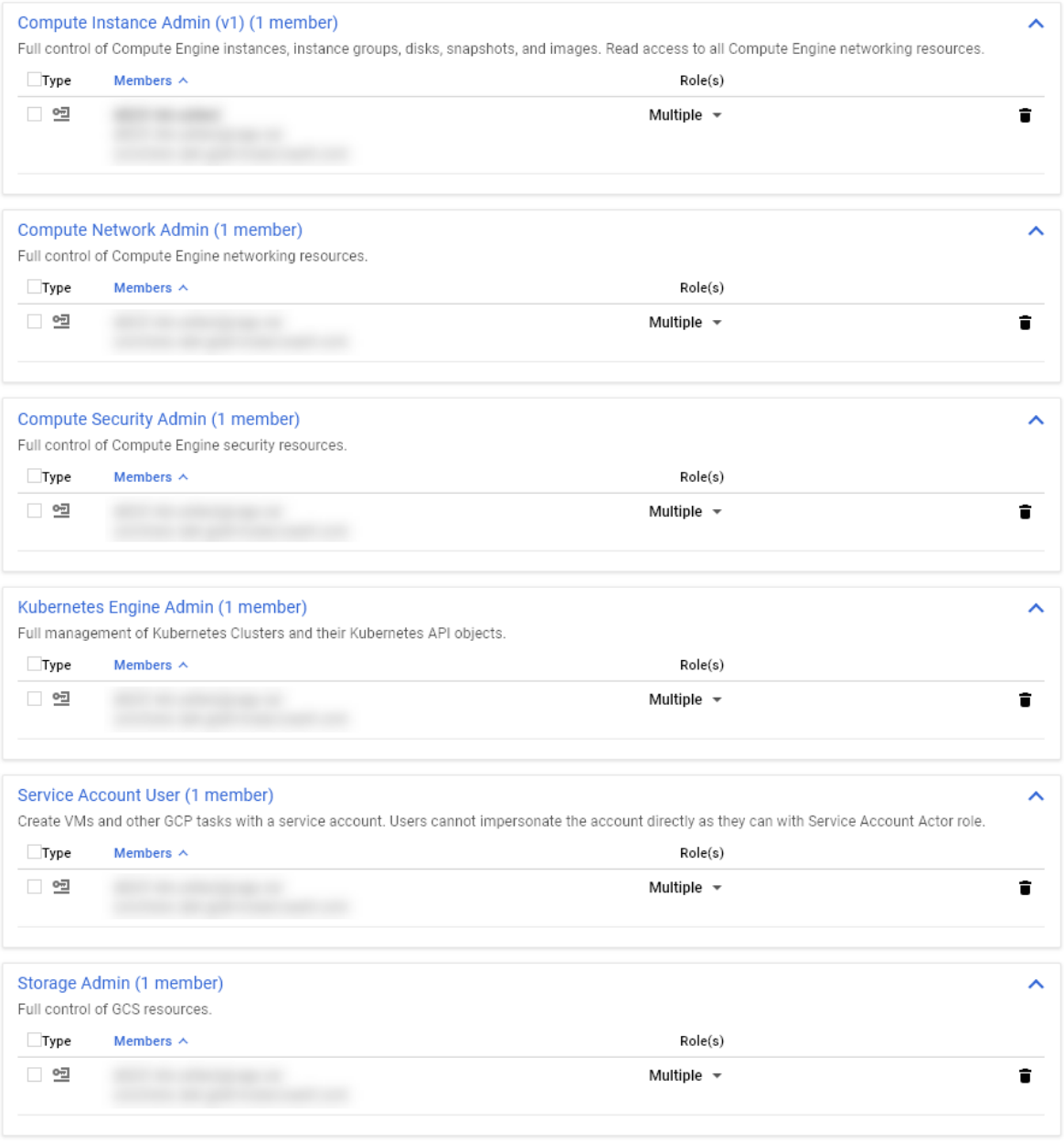
To use the SAP Cloud Appliance Library in order to create a solution instance of SAP Data Hub, trial edition, you need to have access to a Google Cloud Platform project through a service account.

You can grant roles to a service account to ensure that it has permission to complete specific actions on the resources of the Google Cloud Platform project. Please ensure that the service account you use has 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 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.

Attention: If you don't find the roles (in particular Cloud Resource Manager API, Compute Engine API and Kubernetes Engine API), then activate the corresponding APIs and services inside Google Cloud Platform first. Refer to the Google Cloud Documentation for details.

In the Google Cloud Platform Console the service account will look like this.



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

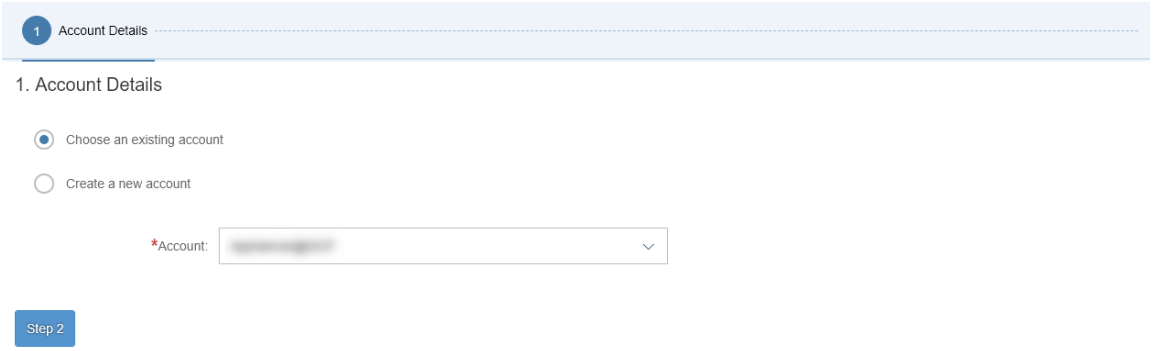
Open the SAP Cloud Appliance Library in your web browser using the following link: https://cal.sap.com.

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.

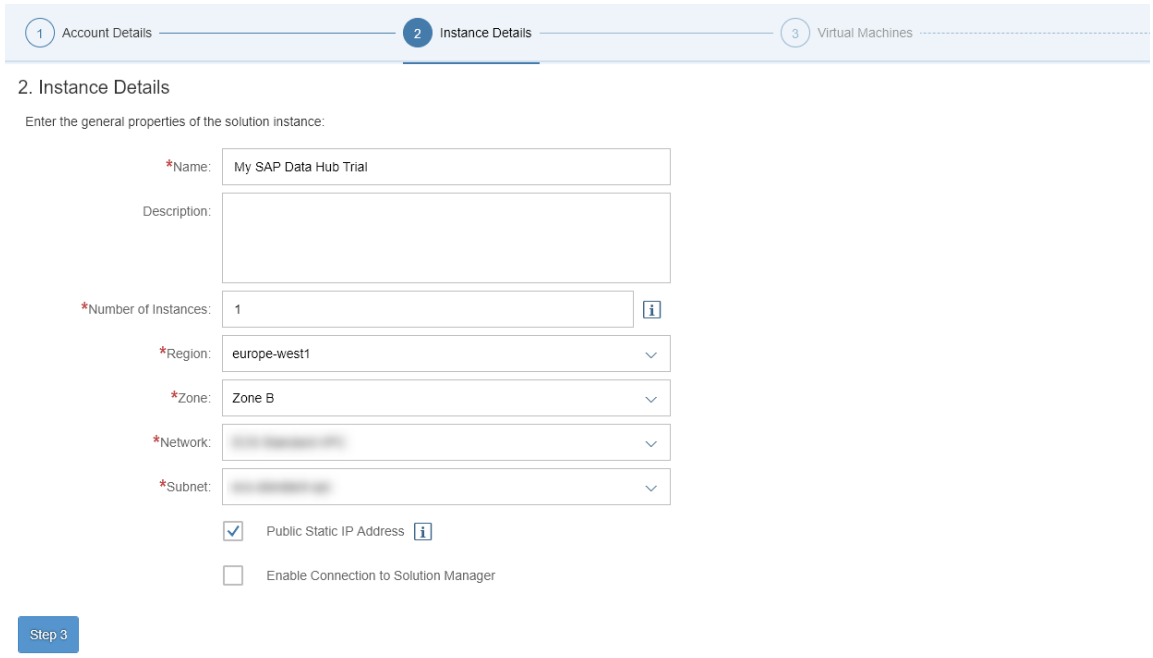
2.3 Creating a Solution Instance

The following steps guide you through the creation of a solution instance of SAP Data Hub 1.0 SPS 04, trial edition. Wherever you need more information, refer to the documentation.

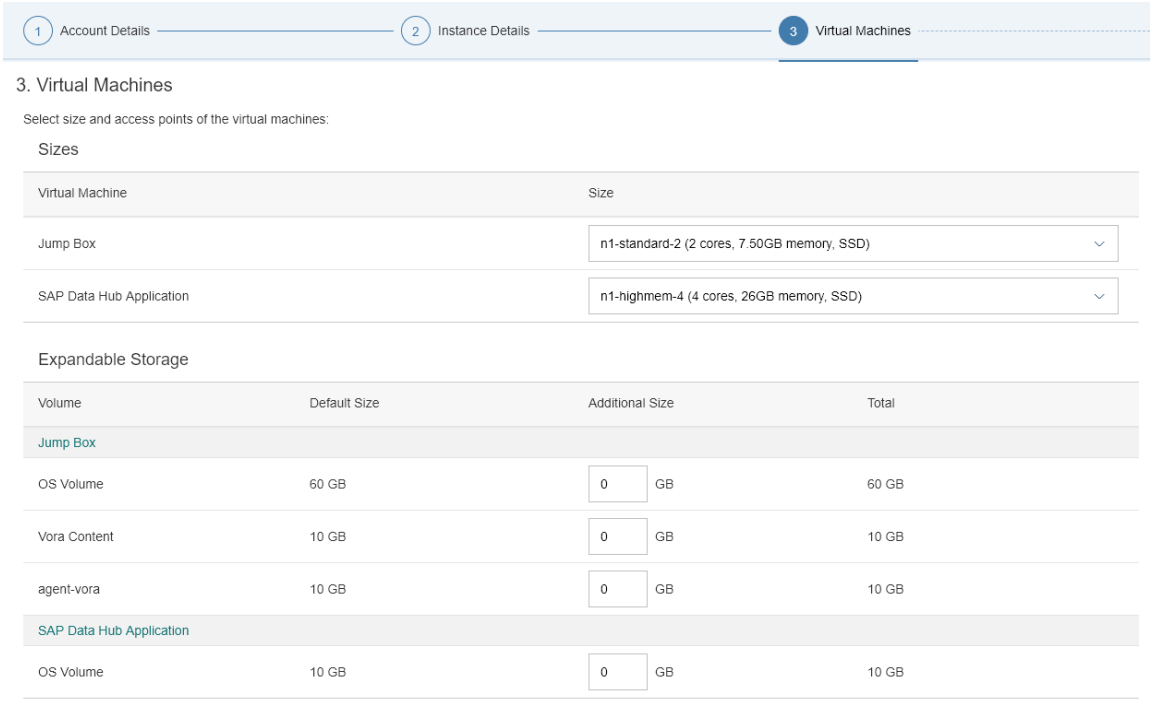
1. Navigate to solution SAP Data Hub 1.0 SPS 04, trial edition.
2. Press the Create Instance button and switch to Advanced Mode (pressing the button in the lower right of the screen).
3. Select the cloud provider account.



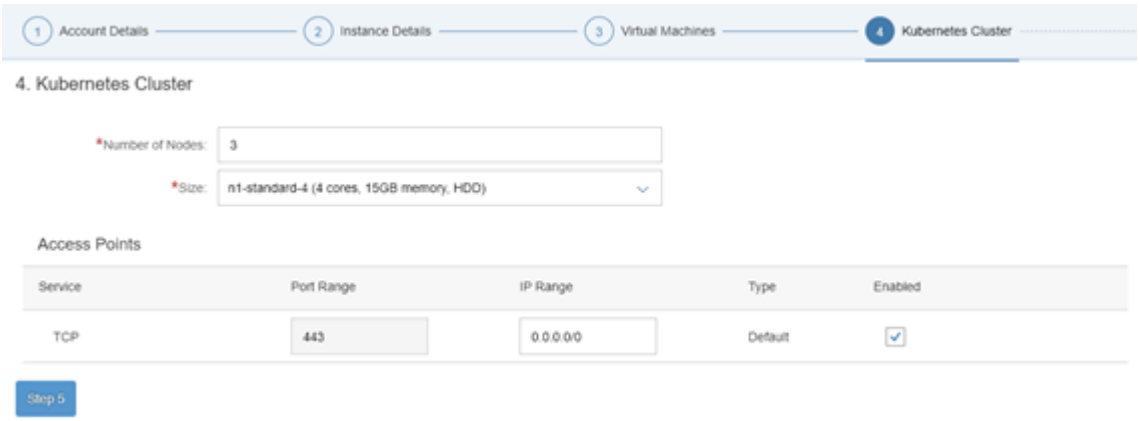
4. Enter Name, Number of Instances (1), Region, Zone, Network and Subnet. Ensure that the checkbox Public Static IP Address is marked.



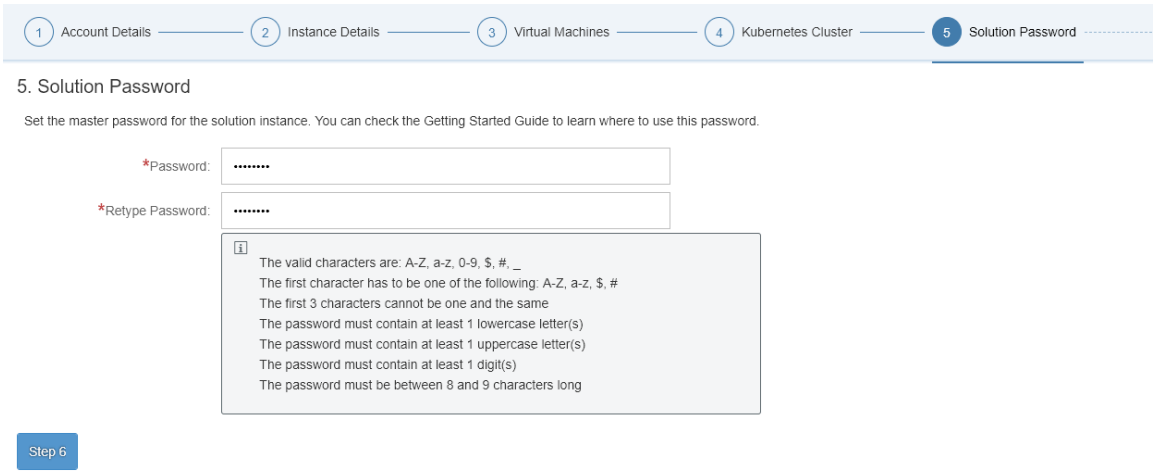
5. Next check the parameters of the Jump Box and the SAP Data Hub Application: Confirm the size of the virtual machines, the expandable storage and the access points. Normally you do not have to change anything.



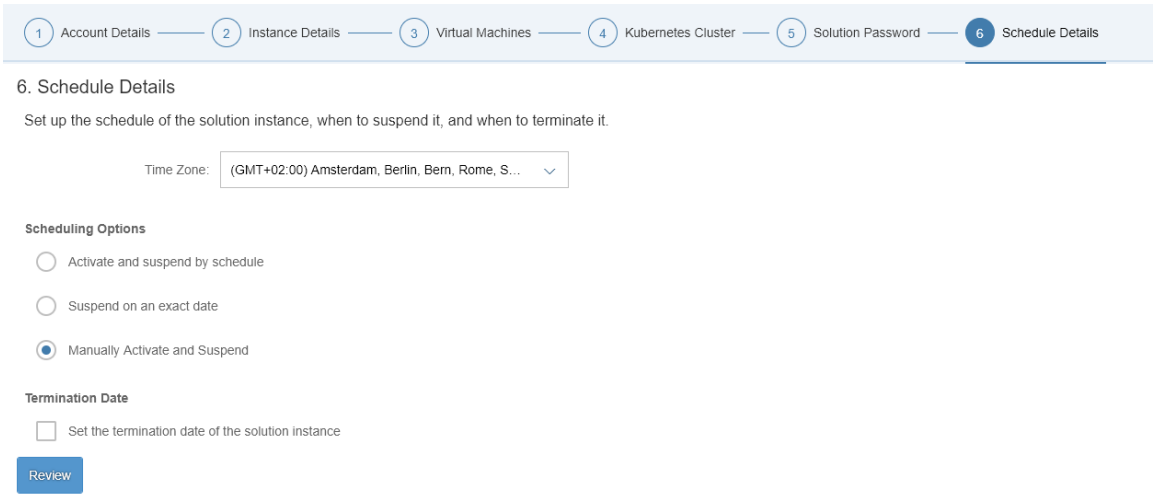
6. Next check the parameters of the SAP Data Hub Distributed Runtime: Confirm the number of nodes, their size and the access points. Normally you do not have to change anything.



7. Next define a solution password. We will refer to this as <Master Password> for the rest of this document.
8. Finally set up the schedule of the solution instance. You can define a schedule to suspend it. And you can



set a termination date.



9. Press the Review button. Then (assuming you are happy with all parameters) press the Create button. It will now take around 40 minutes to create the solution instance.

   You can store and/or download the private key. This is needed in case you want to access the SAP Data Hub Application or the Jump Box on operating system level (typically only needed for troubleshooting).

3 Accessing the Solution

3.1 Overview

As already described in chapter 1.2 Architecture Overview, a solution instance of SAP Data Hub, trial edition consists of:

- a virtual machine to orchestrate the solution instance creation (Jump Box)
- a virtual machine to run the SAP Data Hub Application
- a Kubernetes cluster to run the SAP Data Hub Distributed Runtime

Subsequently you learn how to access your solution instance. A precondition is that you map certain IP addresses against hostnames. This is described in chapter 3.2 Mapping your local hosts file.

Afterwards you can proceed to access the different parts of the solution instance as described in:

- Chapter 3.3 Accessing the SAP Data Hub Application
- Chapter 3.4 Accessing the SAP Data Hub Distributed Runtime

For troubleshooting you might also want to access the Jump Box. You find a description for this in chapter 4 Connecting to Google Cloud Storage

While evaluating and testing SAP Data Hub, you often want to connect to a data lake and/or object storage. With SAP Data Hub, you can connect to Amazon S3, Azure Data Lake (ADL), Azure Storage Blobs (WASB), HDFS and Google Cloud Storage.

Since the SAP Data Hub, trial edition uses Google Cloud Platform as cloud provider, you can easily use Google Cloud Storage (connections to other object storages are technically possible as well):

First create a bucket in Google Cloud Storage (see Google Cloud Documentation). We strongly recommend creating the bucket in the same region you have deployed the SAP Data Hub, trial edition to. This avoids network and latency problems.
For the sake of simplicity, you can create the bucket in the same Google Cloud Platform project where the solution instance was created (but that is not a must).

1. Then open the SAP Data Hub Cockpit (https://vhcalhxedb:51076/) with the user DATAHUB and password <Master Password> (see 3.3.1 Accessing the Application UI).
2. Navigate to Landscape Management and create a new connection by pressing the Add button (in the lower right of the screen).
3. Use GCS_CONN_DEFAULT as ID for the connection (the ID matters!). Then enter the connection details. If you created the bucket in the same Google Cloud Platform project where the solution instance was created, you are able to use the same service account (see 2.1 Prerequisite: Cloud Provider Account) and hence the same key file as when linking the SAP Cloud Appliance Library with the Google Cloud Platform project.
4. Validate the connection. Then (assuming the validation was successful) save the connection by pressing the Add button.
5. Now edit the (already existing) VORA_CATALOG_CONN_DEFAULT connection by pressing the Edit Connection button (in the upper right of the screen).

6. Use GCS as storage type and the bucket you have created to enter the URI (the format is gs://<Bucket Name>). Confirm the password (this is the <Master Password>).
   Validate the connection. Then (assuming the validation was successful) save the connection by pressing the Save button.

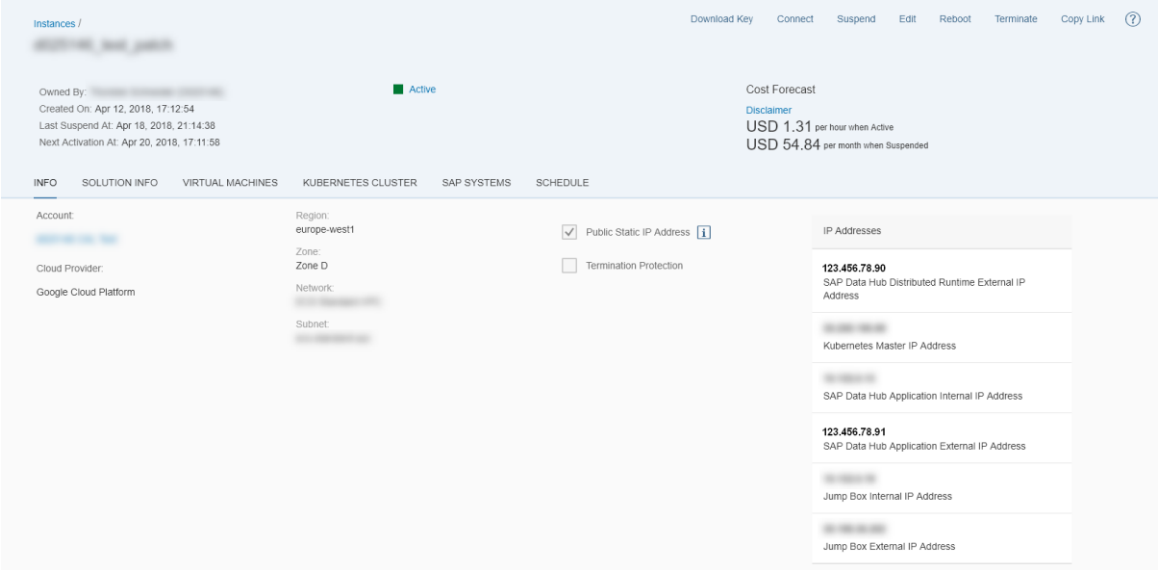
Accessing the Jump Box.

Remark: This chapter is not a step-by-step description how to work with SAP Data Hub (for this you can refer to the links provided in chapter 1.1 Product Overview). It is rather intended as reference. As end-user, you can navigate to all relevant tools / user interfaces directly from the SAP Data Hub Cockpit. This is described in chapter 3.3.1 Accessing the Application UI. Make use of this possibility.

3.2 Mapping your local hosts file

All solution instances of SAP Data Hub, trial edition use the same hostnames vhcalhxedb and vhcalruntime. These hostnames are not fully qualified and you hence need to map the actual (external) IP addresses against them. By doing so, your local computer can resolve the hostnames whenever you use them (for example to access the SAP Data Hub Application via a web browser). Proceed as follows to do the mapping:

1. Look up the external IP addresses for your solution instance in the SAP Cloud Appliance Library. Example (for your solution instance the IP addresses will be different!):
   - SAP Data Hub Application External IP Address: 123.456.78.91
   - SAP Data Hub Distributed Runtime External IP Address: 123.456.78.90



2. Open a text editor (e.g. Notepad in case of Microsoft Windows) as administrator (e.g. for Microsoft Windows search for Notepad on your computer, open the context menu of Notepad and click Run as administrator).
3. Click File → Open and enter the following path:
   - For Microsoft Windows operating system: c:\windows\system32\drivers\etc\hosts
   - For Linux operating system: /etc/hosts
4. Ensure to select All Files (*.*).
5. Open the hosts file and add the following lines to it:

```
# SAP Data Hub, trial edition
# SAP Data Hub Application
<SAP Data Hub Application External IP Address> vhcalhxedb
# SAP Data Hub Distributed Runtime
<SAP Data Hub Distributed Runtime External IP Address> vhcalruntime
```

6. Replace <SAP Data Hub Application External IP Address> and <SAP Data Hub Distributed Runtime External IP Address> by the external IP addresses which you have retrieved from SAP Cloud Appliance Library. Example:

```
# SAP Data Hub, trial edition
# SAP Data Hub Application
123.456.78.91 vhcalhxedb
# SAP Data Hub Distributed Runtime
123.456.78.90 vhcalruntime
```

7. Save the hosts file and exit the text editor.

3.3 Accessing the SAP Data Hub Application

3.3.1 Accessing the Application UI

The Application UI particularly consists of the SAP Data Hub Cockpit and the SAP Data Hub Modeling tool:

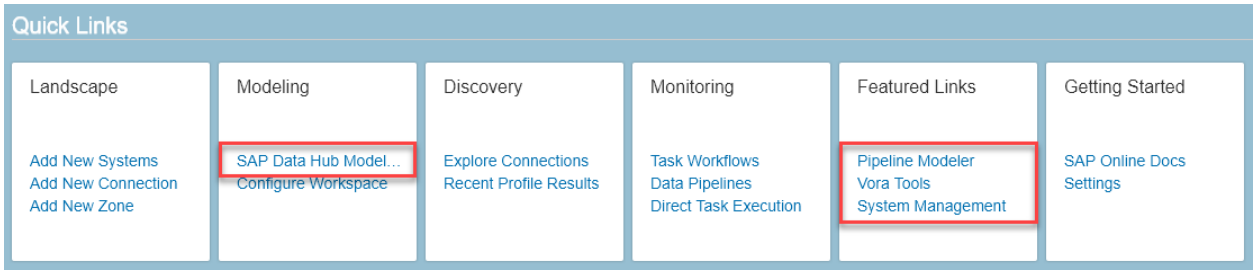
- The SAP Data Hub Cockpit is the central entry point for end-users of SAP Data Hub. It provides you with a single point of access to a range of tools.
- The SAP Data Hub Modeling tool allows you to create task workflows.

You can access the Application UI via a web browser (Microsoft Internet Explorer, Google Chrome, Mozilla Firefox) using the information in the tables below:

SAP Data Hub Cockpit

Name	Value	Description
URL	https://vhcalhxedb:51076/	SAP Data Hub Cockpit
User	DATAHUB	User for the Application UI
Password	<Master Password>	The initial master password of the system you provided in the SAP Cloud Appliance Library when creating the instance.

Attention: If you get an error" There is no any resources matched to request path /extension/bdh/tools/overview" while opening the SAP Data Hub Cockpit, then please proceed as described in 7 Troubleshooting to resolve this problem.



From the SAP Data Hub Cockpit, you can use the Quick Links to navigate to the SAP Data Hub Modeling tool as SAP Data Hub System Management, SAP Data Hub Pipeline Modeler as well as SAP Vora Tools. For the three last-mentioned you need to provide tenant, user and password after clicking on the link. You find this information in chapter 3.3.4 Accessing the Distributed Runtime UIs.

SAP Data Hub Modeling tool

Name	Value	Description
URL	https://vhcalhxedb:51058/	SAP Data Hub Modeling tool
User	DATAHUB	User for the Application UI
Password	<Master Password>	The initial master password of the system you provided in the SAP Cloud Appliance Library when creating the instance.

Remark: You might see an error message that the web browser is not able to verify the security certificate of https://vhcalhxedb. The reason is that the used security certificate is self-signed and the web browser therefore cannot find a trusted root certificate. You need to confirm the error message to proceed. Look at chapter 7 Troubleshooting in case of problems.

3.3.2 Accessing the XS Advanced Administration tool

The XS Advanced Administration tool, for example, allows you to manage application roles and users, monitor applications and view audit logs.

You can access the XS Advanced Administration tool via a web browser (Microsoft Internet Explorer, Google Chrome, Mozilla Firefox) using the information in the table below:

Name	Value	Description
URL	https://vhcalhxedb:51015/	XS Advanced Administration tool
User	XSA_ADMIN	User for the XSA Administration and Monitoring Tools
Password	<Master Password>	The initial master password of the system you provided in the SAP Cloud Appliance Library when creating the instance.

Remark: You might see an error message that the web browser is not able to verify the security certificate of https://vhcalhxedb. The reason is that the used security certificate is self-signed and the web browser therefore cannot find a trusted root certificate. You need to confirm the error message to proceed.

3.3.3 Accessing the SAP HANA database

The SAP Data Hub Application uses the SAP HANA database as persistence. All data is stored in a dedicated multi-tenant database container HXE. You can access the SAP HANA database either via SAP HANA Cockpit or SAP HANA Tools (Eclipse plug-in).

SAP HANA Cockpit

Name	Value	Description
URL	https://vhcalhxedb:51041	SAP HANA Cockpit
User	XSA_ADMIN	User for the XSA Administration and Monitoring Tools
Password	<Master Password>	The initial master password of the system you provided in the SAP Cloud Appliance Library when creating the instance.

Remark: You might see an error message that the web browser is not able to verify the security certificate of https://vhcalhxedb. The reason is that the used security certificate is self-signed and the web browser therefore cannot find a trusted root certificate. You need to confirm the error message to proceed.

SAP HANA Tools

Name	Value	Description
Hostname	vhcalhxedb	Hostname of the SAP HANA system
Instance Number	90	Instance number of the central instance of the SAP System
Mode	Multiple Containers	The SAP HANA system is configured for multi-tenant database containers.
Database	System database Tenant database HXE	You can connect to both, the system database as well as tenant database HXE. Attention: To connect to tenant database HXE, you need to follow the description at the end of this chapter prior to creating the connection.
Username	SYSTEM XSA_ADMIN	These are the standard users which you can use to access the database server.
Password	<Master Password>	The initial master password of the system you provided in the SAP Cloud Appliance Library when creating the instance.

Connecting from SAP HANA Tools to tenant database HXE:

To connect from SAP HANA Tools to tenant database HXE, you once need to configure hostname resolution prior to creating the connection. Thereto proceed as follows in SAP HANA Studio:

1. Connect to the system database using the SYSTEM user.
2. Open the Administration.
3. Click on the Configuration tab and use the value use_default_route as a filter.

4. Open the context menu of public_hostname_resolution and click Add parameter....
5. Select System as scope. Press Next.
6. Enter map_vhcalhxedb as key and <SAP Data Hub Application External IP Address> (i.e. the external IP address which you have also maintained in the local hosts file) as value.

3.3.4 Accessing the Distributed Runtime UIs

The Distributed Runtime UIs consist of the SAP Data Hub System Management, the SAP Data Hub Pipeline Modeler and the SAP Vora Tools:

- The SAP Data Hub System Management allows you to manage applications, in particular the SAP Data Hub Pipeline Modeler and the SAP Vora Tools, as well as corresponding users.
- The SAP Data Hub Pipeline Modeler allows you to create data-driven applications, so-called data pipelines.
- The SAP Vora Tools provide you with a data modeling environment for creating and maintaining tables and views.

You can access the Distributed Runtime UIs via a web browser (Microsoft Internet Explorer, Google Chrome, Mozilla Firefox) using the information in the tables below:

SAP Data Hub System Management

Name	Value	Description
URL	https://vhcalruntime/home/	SAP Data Hub System Management
Tenant ID	system default	System tenant Default tenant Attention: SAP Data Hub Pipeline Modeler and SAP Vora Tools are only available in default tenant.
User	DATAHUB	User for the Distributed Runtime UIs
Password	<Master Password>	The initial master password of the system you provided in the SAP Cloud Appliance Library when creating the instance.

SAP Data Hub Pipeline Modeler

Name	Value	Description
URL	https://vhcalruntime/app/pipeline-modeler	SAP Data Hub Pipeline Modeller
Tenant ID	Default	Default tenant Attention: if you are logged into the system tenant (in another browser window) SAP Data Hub Pipeline Modeler is not available (error 404). Log off from the system tenant.
User	DATAHUB	User for the Distributed Runtime UIs
Password	<Master Password>	The initial master password of the system you provided in the SAP Cloud Appliance Library when creating the instance.

SAP Vora Tools

Name	Value	Description
URL	https://vhcalruntime/app/vora-tools	SAP Vora Tools
Tenant ID	default	Default tenant Attention: if you are logged into the system tenant (in another browser window) SAP Vora Tools is not available (error 404). Log off from the system tenant.
User	DATAHUB	User for the Distributed Runtime UIs
Password	<Master Password>	The initial master password of the system you provided in the SAP Cloud Appliance Library when creating the instance.

Remark: You might see an error message that the web browser is not able to verify the security certificate of https://vhcalruntime. The reason is that the used security certificate is self-signed and the web browser therefore cannot find a trusted root certificate. You need to confirm the error message to proceed.

3.3.5 Accessing the Operating System

You can access the SAP Data Hub Application on operating system level via the secure shell protocol. 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	root / super user with access to all commands and files
hxeadm	SAP system administrator
hxeshm	Group needed by SAP HANA database on OS level for shared memory operations
sapadm	SAP database administrator
sapsys	Group containing all <SID>adm users (should be a group in a central user storage like LDAP, NIS, or Active Directory)

3.4 Accessing the SAP Data Hub Distributed Runtime

3.4.1 Accessing the Kubernetes Cluster

The SAP Data Hub Distributed Runtime uses Kubernetes. In case of the SAP Data Hub, trial edition it uses the Google Kubernetes Engine.

You can access the Kubernetes cluster used by the SAP Data Hub, trial edition via the Google Cloud Platform Console as well as the Google Cloud SDK (in particular gcloud and kubectl).

3.4.2 Accessing SAP Vora Diagnostics UIs

SAP Vora Diagnostics is an open-source toolchain that supports you with monitoring and troubleshooting. It provides you with Grafana (for metrics monitoring) and Kibana (for trace log analysis) UIs.

You can access the SAP Vora Diagnostics UIs via a web browser (Microsoft Internet Explorer, Google Chrome, Mozilla Firefox). Thereto you need to create port forwards as described in the documentation. This requires that you have the Google Cloud SDK installed on your local computer.

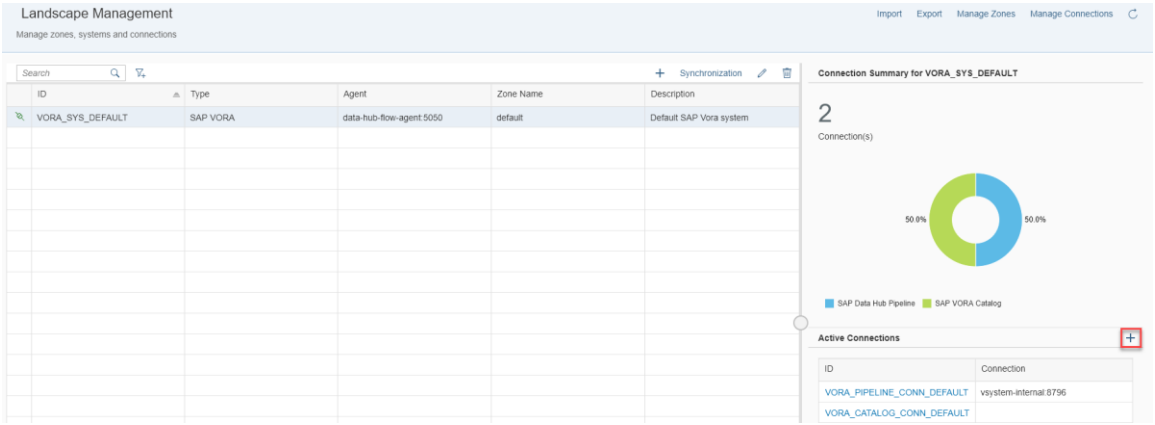
4 Connecting to Google Cloud Storage

While evaluating and testing SAP Data Hub, you often want to connect to a data lake and/or object storage. With SAP Data Hub, you can connect to Amazon S3, Azure Data Lake (ADL), Azure Storage Blobs (WASB), HDFS and Google Cloud Storage.

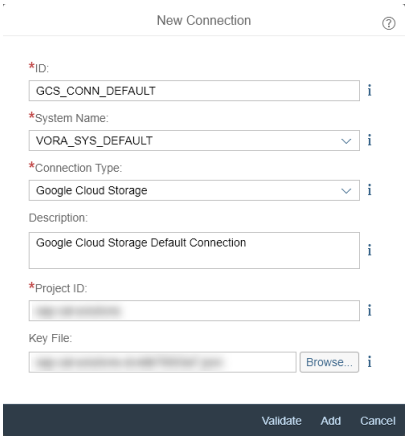
Since the SAP Data Hub, trial edition uses Google Cloud Platform as cloud provider, you can easily use Google Cloud Storage (connections to other object storages are technically possible as well):

First create a bucket in Google Cloud Storage (see Google Cloud Documentation). We strongly recommend creating the bucket in the same region you have deployed the SAP Data Hub, trial edition to. This avoids network and latency problems.
For the sake of simplicity, you can create the bucket in the same Google Cloud Platform project where the solution instance was created (but that is not a must).

7. Then open the SAP Data Hub Cockpit (https://vhcalhxedb:51076/) with the user DATAHUB and password <Master Password> (see 3.3.1 Accessing the Application UI).
8. Navigate to Landscape Management and create a new connection by pressing the Add button (in the lower right of the screen).

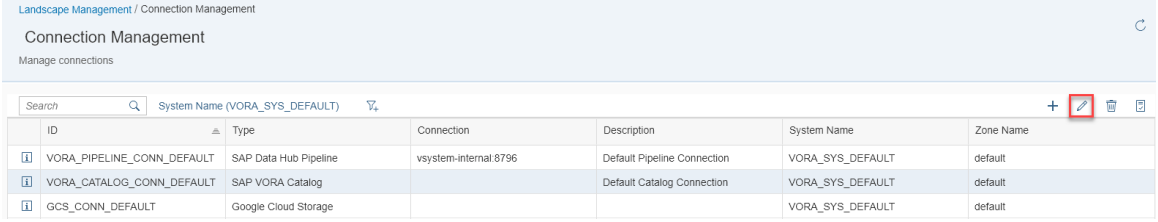


9. Use GCS_CONN_DEFAULT as ID for the connection (the ID matters!). Then enter the connection details.

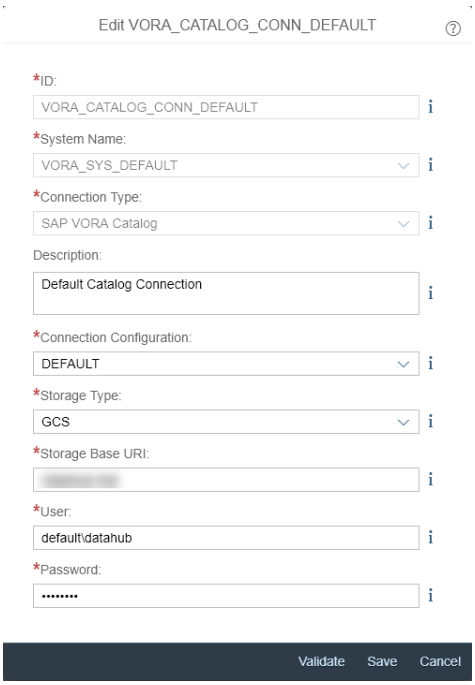


If you created the bucket in the same Google Cloud Platform project where the solution instance was created, you are able to use the same service account (see 2.1 Prerequisite: Cloud Provider Account) and hence the same key file as when linking the SAP Cloud Appliance Library with the Google Cloud Platform project.

10. Validate the connection. Then (assuming the validation was successful) save the connection by pressing the Add button.
11. Now edit the (already existing) VORA_CATALOG_CONN_DEFAULT connection by pressing the Edit Connection button (in the upper right of the screen).



12. Use GCS as storage type and the bucket you have created to enter the URI (the format is gs://<Bucket Name>). Confirm the password (this is the <Master Password>).



Validate the connection. Then (assuming the validation was successful) save the connection by pressing the Save button.

4.1 Accessing the Jump Box

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	root / super user with access to all commands and files

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 the SAP Data Hub Distributed Runtime.

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

For more information refer to the Kubernetes documention.

5 Licenses

5.1 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.

5.2 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 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.

6 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 Linux-based servers
HTTP(s)*	4390	SAP Web Dispatcher (HANA)
HTTP(s)*	8090	SAP Web Dispatcher (HANA)
Custom*	39013	SQL and MDX access port to the SYSTEM database
Custom*	39015	SQL and MDX access to the first tenant of a HANA system – in the case of HXE, a user has to initiate the tenant before trying to connect to this port
Custom*	39041-39045	Additional open ports for 5 Tenants
Custom*	39017	Port for statistics server connections
HTTP(s)*	39033	Default XSA port if hostname routing is used for HANA
HTTP(s)*	39032	Default XSA port if hostname routing is used for HANA
HTTP(s)*	39030	XS-controller managed Web Dispatcher
HTTP(s)*	59013	Instance Agent
HTTP(s)*	59014	Instance Agent SSL
HTTP(s)	51000-51099	XSA application instances
Custom*	39026-39030	Ports for SDS streaming clients
Custom*	39040	Default DPServer Port
Custom*	1128-1129	SAP Host Agent
Custom*	53075	WEB based IDE for XSA Development
Custom*	53030	DICore Service
TCP	443	SAP Data Hub System Management

* from solution “SAP HANA, express edition”

You must change the initial user passwords provided by SAP when you log onto the system for the first time.

Note that when using **HANA based appliances**, HANA systems are not installed individually but **cloned from a template system**. As a consequence of this cloning process, the existing root keys are cloned. For more information, see this SAP Note 2134846 - HANA encryption key handling during system cloning.

For more information about security vulnerabilities, see this community page.

7 Troubleshooting

Error “There is no any resources matched to request path /extension/bdh/tools/overview”

You open the SAP Data Hub Cockpit. The link to the SAP Data Hub Modeling tool is not available.
You see an error message at the bottom of the screen “There is no any resources matched to request path /extension/bdh/tools/overview”.

Reason: program error (will be fixed in a future version of SAP Data Hub, trial edition).

Solution: Access the operating system of the virtual machine (hostname vhcalhxedb) hosting the SAP Data Hub Application as described in 3.3.5 Accessing the Operating System and run the following commands.
Replace <Master Password> with the password you provided in the SAP Cloud Appliance Library when creating the instance.

```
sudo su - hxeadm
xs login -a https://vhcalhxedb:39030 -u XSA_ADMIN -p <Master Password> -o
HANAExpress -s DATAHUB --skip-ssl-validation
xs restart webide
xs restart di-core
```

Error during profiling “Adapter Error Message “code”: “ECONNRESET”

You profile a file in Discovery. After a while the file’s profiling status is displayed as erroneous. In SAP Data Hub Pipeline Modeler you see a PROFILE::com.sap.data.discovery.generated.objects... pipeline. This pipeline initially in status pending and later in status dead.

Reason: timeout (system behavior will be improved)

Solution: Wait around 5 minutes (more precisely until the PROFILE::com.sap.data.discovery.generated.objects... pipeline is dead). Then start profiling again. Typically, that will solve the problem once and for all (and the system will display the file’s profiling status as “Profiled”). If it does not, start the profiling a third time.

Error “404 page not found“ when trying to open SAP Data Hub Pipeline Modeler or SAP Vora Tools

You try to open the SAP Data Hub Pipeline Modeler or the SAP Vora Tools. The web browser shows the error “404 page not found”. This happens independent of whether you try to open the aforementioned UIs from the SAP Data Hub Cockpit or directly.

Reason: You are logged into the system tenant. Neither the SAP Data Hub Pipeline Modeler nor the SAP Vora Tools are available in this tenant.

Solution: Log off from the system tenant. Use the default tenant instead.

SAP Data Hub Application UIs cannot be accessed

You open the SAP Data Hub Cockpit and the screen "stays blue". You open the SAP Data Hub Modeling tool and certain plug-ins cannot be loaded.

Reason: Most likely you use Microsoft Internet Explorer or Mozilla Firefox and the security settings you have configured for these web browsers do not allow them to load (certain) UI resources.

Solution: You need to ensure that the web browser of your choice trusts the self-signed certificate used by SAP Data Hub, trial edition and can load all needed UI resources. For that consult the documentation of the respective web browser.

The following approach might help you: Determine via the network trace of your web browser, which UI resources cannot be loaded (e.g. https://vhcalhxedb:51035/resources/sap-ui-core.js). Then open the URL (e.g. https://vhcalhxedb:51035/resources/sap-ui-core.js) directly in the web browser of your choice and confirm the certificate / site exception. Afterwards try again to open the SAP Data Hub Cockpit respectively the SAP Data Hub modeling tool.

Error "Gateway Timeout" in SAP Data Hub Cockpit

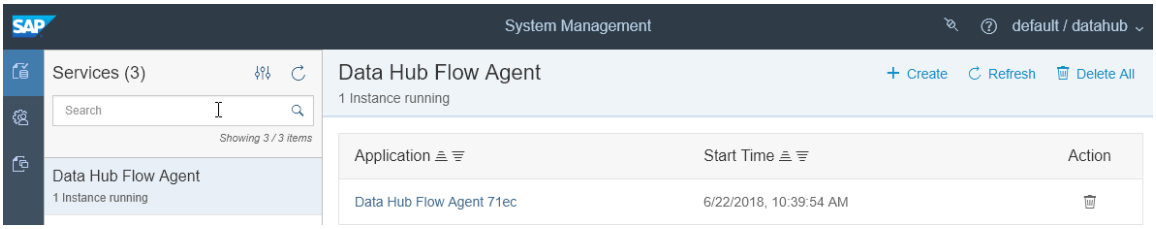
You open the SAP Data Hub Cockpit. The tiles visualizing the system status are not loaded and you get an error "Gateway Timeout".

Reason: The error is typically caused by connectivity problems between the SAP Data Hub Application and the SAP Data Hub Distributed Runtime. Connectivity problems can have many reasons (which under "normal" circumstances will not occur in the SAP Data Hub, trial edition). One reason for connectivity problems can be that you deleted the instance of the SAP Data Hub Flow Agent which was started while creating the solution instance of SAP Data Hub, trial edition.

Even if you afterwards create a new instance of SAP Data Hub Flow Agent, the internal load balancer used for the communication between the SAP Data Hub Application and the SAP Data Hub Distributed Runtime will not point to the correct Kubernetes service / pod. Hence avoid deleting the instance of SAP Data Hub Flow Agent.

Solution: First, wait 10-15 minutes (to avoid that a "general" network problem on Google Cloud Platform is the reason for the error). If waiting does not help, restart the solution instance of SAP Data Hub, trial edition (by suspending and then activating it again). Only if restarting does not help, proceed as follows:

1. Open SAP Data Hub System Management (https://vhcalruntime/home/) for the default tenant and ensure that exactly one instance of SAP Data Hub Flow Agent is running. If there is no instance running, create one by pressing the Create button.



2. Afterwards log onto the Jump Box on operating system level via the secure shell protocol (see chapter 4 Connecting to Google Cloud Storage

While evaluating and testing SAP Data Hub, you often want to connect to a data lake and/or object storage. With SAP Data Hub, you can connect to Amazon S3, Azure Data Lake (ADL), Azure Storage Blobs (WASB), HDFS and Google Cloud Storage.

Since the SAP Data Hub, trial edition uses Google Cloud Platform as cloud provider, you can easily use Google Cloud Storage (connections to other object storages are technically possible as well):

First create a bucket in Google Cloud Storage (see Google Cloud Documentation). We strongly recommend creating the bucket in the same region you have deployed the SAP Data Hub, trial edition to. This avoids network and latency problems.
For the sake of simplicity, you can create the bucket in the same Google Cloud Platform project where the solution instance was created (but that is not a must).

13. Then open the SAP Data Hub Cockpit (https://vhcalhxedb:51076/) with the user DATAHUB and password <Master Password> (see 3.3.1 Accessing the Application UI).
14. Navigate to Landscape Management and create a new connection by pressing the Add button (in the lower right of the screen).
15. Use GCS_CONN_DEFAULT as ID for the connection (the ID matters!). Then enter the connection details. If you created the bucket in the same Google Cloud Platform project where the solution instance was created, you are able to use the same service account (see 2.1 Prerequisite: Cloud Provider Account) and hence the same key file as when linking the SAP Cloud Appliance Library with the Google Cloud Platform project.
16. Validate the connection. Then (assuming the validation was successful) save the connection by pressing the Add button.
17. Now edit the (already existing) VORA_CATALOG_CONN_DEFAULT connection by pressing the Edit Connection button (in the upper right of the screen).
18. Use GCS as storage type and the bucket you have created to enter the URI (the format is gs://<Bucket Name>). Confirm the password (this is the <Master Password>).
    Validate the connection. Then (assuming the validation was successful) save the connection by pressing the Save button.
3. Accessing the Jump Box).
4. Run the following command to confirm that the Kubernetes pod for SAP Data Hub Flow Agent is running.

```
kubectl --namespace datahub get pods | grep data-hub-flow-agent
```

The command will return an output similar to the following.

5. Run the following commands to determine the APP_SELECTOR.

```
export SERVICE=`kubectl --namespace datahub get services| grep data-hub-flow-agent | grep ClusterIP`
echo ${SERVICE}
export SERVICE_NAME=`echo ${SERVICE} | cut -d " " -f1`
echo ${SERVICE_NAME}
export APP_SELECTOR=`kubectl -n datahub get service ${SERVICE_NAME} --output jsonpath='{.spec.selector.app}'`
echo ${APP_SELECTOR}
```

The commands will return an output similar to the following. Write down the value of APP_SELECTOR (e.g. 71eccc3772fd4023a414011b5dce856e).

```
caltdc-50125122-[redacted]-jump-box:~ # export SERVICE=`kubectl --namespace datahub get services| grep data-hub-flow-agent | grep ClusterIP`
caltdc-50125122-[redacted]-jump-box:~ # echo ${SERVICE}
data-hub-flow-agent-xd241 ClusterIP 10.39.240.113 <none> 5050/TCP 23h
caltdc-50125122-[redacted]-jump-box:~ # export SERVICE_NAME=`echo ${SERVICE} | cut -d " " -f1`
caltdc-50125122-[redacted]-jump-box:~ # echo ${SERVICE_NAME}
data-hub-flow-agent-xd241
caltdc-50125122-[redacted]-jump-box:~ # export APP_SELECTOR=`kubectl -n datahub get service ${SERVICE_NAME} --output jsonpath='{.spec.selector.app}'`
caltdc-50125122-[redacted]-jump-box:~ # echo ${APP_SELECTOR}
71eccc3772fd4023a414011b5dce856e
caltdc-50125122-[redacted]-jump-box:~ #
```

6. Run the following commands to edit the internallb.yaml file.

```
cd /root/postconfig/
vi internallb.yaml
```

Put the name of the APP_SELECTOR as spec → selector → app (see screenshot) and save the file.

```
apiVersion: v1
kind: Service
metadata:
  name: data-hub-flow-agent
  namespace: datahub
  annotations:
    cloud.google.com/load-balancer-type: "Internal"
    cal-cluster: caltdc-50125123-i052582-cluster
spec:
  type: LoadBalancer
  externalTrafficPolicy: Cluster
  loadBalancerIP: 10.132.0.4
  ports:
  - port: 5050
    protocol: TCP
    targetPort: 5050
  selector:
    app: 71eccc3772fd4023a414011b5dce856e
  sessionAffinity: None
```

7. Run the following command type apply the changes to the internal load balancer. This will ensure that the internal load balancer points to the correct Kubernetes service / pod.

```
kubectl -n datahub apply -f internallb.yaml
```

Blank screen when opening “SAP Data Hub Pipeline Modeler”

You open the SAP Data Hub Pipeline Modeler. The screen is empty except the header line.

Reason: program error (will be fixed in a future version of SAP Data Hub, trial edition).

Solution: Add ?reset to the URL (i.e. https://vhcalruntime/app/pipeline-modeler?reset) and press enter.

8 Appendix

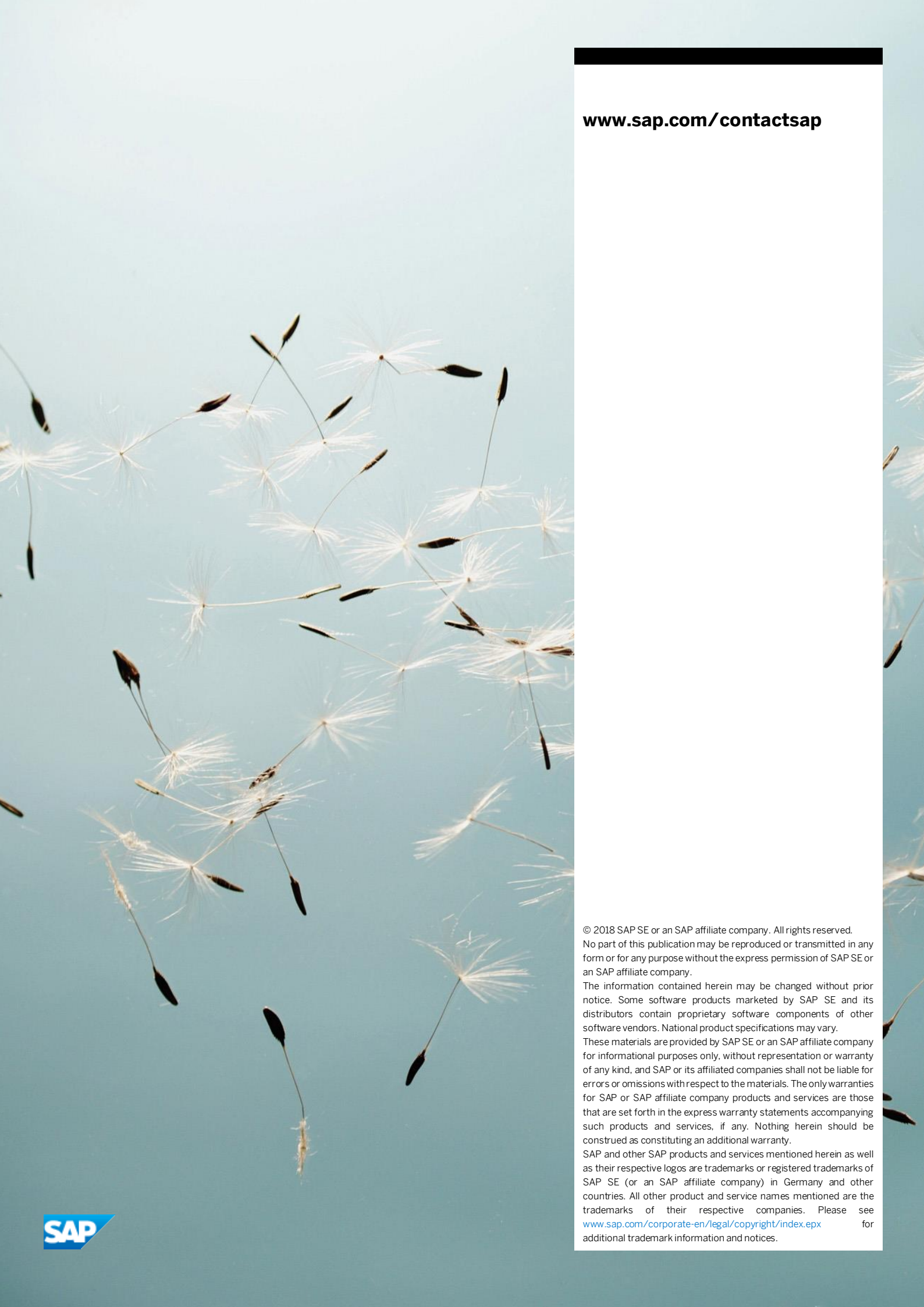
Installed Software Components

Name	Release	Support Package Stack
HDB_LCM_LINUX_X86_64*	2.00.021.00	2
HDB_SERVER_LINUX_X86_64*	2.00.021.00	2
HDB_AFL_LINUX_X86_64*	2.00.021.00	2
HDB_EML_AFL_LINUX_X86_64*	2.00.021.00	2
XSA_RT_10_LINUX_X86_64*	1.0.66	2
XSA_CONTENT_10*	1.0.66	2
XSAC_SAP_WEB_IDE_20	4.2.18 / 4.2.31	2
XSAC_HRTT_20	2.4.65	4
HANA_COCKPIT_20*	2.3.9	2
HCO_HANA_SHINE*	1.202.0	2
SAP_HANA_STREAMING*	2.00.021.00	2
HANA_SDI*	2.00.021.00	2
HANA_DP_AGENT_20_LIN_X86_64 *	2.00.021.00	2
XSA_CLIENT_10*	2.00.021.00	2
HDB_CLIENT_LINUX_X86_64*	2.2.26.1504297370	2
HDB_CLIENT_LINUX_PPC64LE*	2.2.26.1504297370	2
HDB_CLIENT_WINDOWS_X86_64*	2.2.26.1504297370	2
HDB_CLIENT_NTINTEL*	2.2.26.1504297370	2
HDB_CLIENT_MACOS*	2.00.020.00	2
XSAC_DH_COCKPIT_1.0	1.4.5	4
XSAC_DH_DISCOVERY_1.0	1.4.3	4
XSAC_DH_MD_1.0	1.4.2	4
XSAC_DH_OBJECTSERVICES_1.0	1.4.4	4
XSAC_DH_SECURITY_1.0	1.4.4	4
XSAC_DH_TOOLS_1.0	1.4.12	4

Name	Release	Support Package Stack
SAP DATA HUB DISTRIB RUNTM 1.0	2.2.48	4
SAP DATA HUB FLOW AGENT 1.0	1.4.10	4

* from solution “SAP HANA, express edition”

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



www.sap.com/contactsap