

# Getting Started with SAP Data Hub, trial edition 2.3



---

# Table of Contents

<b>1</b>	<b>Solution Information</b> .....	<b>3</b>
1.1	Product Overview .....	3
1.2	Architecture Overview .....	3
1.3	Installed Products .....	4
<b>2</b>	<b>Provisioning the Solution</b> .....	<b>5</b>
2.1	Prerequisite: Cloud Provider Account .....	5
2.1.1	Required APIs.....	5
2.1.2	Service account and roles.....	5
2.2	Accessing the SAP Cloud Appliance Library.....	7
2.3	Creating a Solution Instance .....	8
<b>3</b>	<b>Accessing the Solution</b> .....	<b>11</b>
3.1	Overview.....	11
3.2	Determining the external IP addresses of your solution instance .....	12
3.3	Accessing SAP Data Hub through the Windows Remote Desktop .....	12
3.4	Accessing SAP Data Hub from your local computer .....	15
3.5	Accessing the Jump Box for troubleshooting .....	17
<b>4</b>	<b>Licenses</b> .....	<b>18</b>
4.1	Running Solution with a Product License Key .....	18
4.2	Running Solution as a Trial for a Free Period .....	18
<b>5</b>	<b>Security Aspects</b> .....	<b>19</b>
<b>6</b>	<b>Troubleshooting</b> .....	<b>20</b>
<b>7</b>	<b>Appendix</b> .....	<b>21</b>

---

# 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.
- Data Workflows orchestrate processes across a data landscape.
- And Data Governance allows you to extract metadata from connected data stores. It supports discovery, 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](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)
- a Kubernetes cluster to run SAP Data Hub
- a Windows Remote Desktop from which you can easily connect to SAP Data Hub

The Jump Box 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 SAP Data Hub. 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).

SAP Data Hub 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 and Google Cloud Storage (GCS)) and to run highly scalable data flows and data-driven applications. As part of this pre-configured appliance SAP Data Hub uses a Kubernetes cluster with three worker nodes (each 8 cores, 30 GB RAM).

The Windows Remote Desktop (4 cores, 15 GB RAM) allows you to easily connect to SAP Data Hub without having to maintain your local hosts file.

---

## 1.3 Installed Products

The SAP Data Hub, trial edition leverages release 2.3 (patch 3).

## 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. In the Google Cloud Platform project, you need to enable certain APIs and to create a service account with certain roles.

#### 2.1.1 Required APIs

To create a solution instance of SAP Data Hub, 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.

#### 2.1.2 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 resources.
Compute Network Admin	Full control of Compute Engine networking resources.

Name	Description
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.

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

The screenshot displays six service account role cards in the Google Cloud Platform Console. Each card represents a different role and includes the following information:

- Role Name:** Compute Instance Admin (v1) (1 member), Compute Network Admin (1 member), Compute Security Admin (1 member), Kubernetes Engine Admin (1 member), Service Account User (1 member), and Storage Admin (1 member).
- Description:** A brief description of the role's permissions, such as "Full control of Compute Engine instances, instance groups, disks, snapshots, and images." for Compute Instance Admin (v1).
- Members:** A link to view the members of the role.
- Table:** A table with columns for 'Type', 'Role(s)', and a trash icon. The 'Role(s)' column shows 'Multiple' for each role.

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

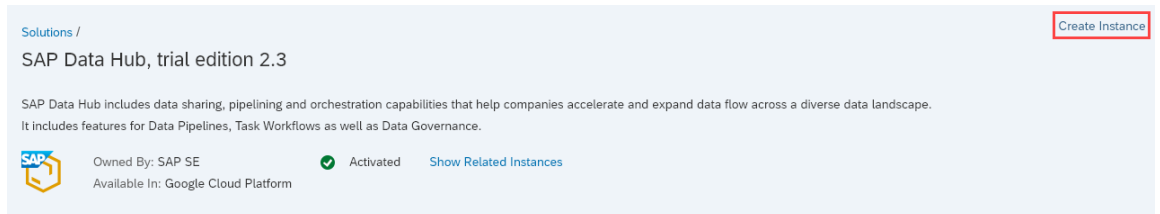
Open the SAP Cloud Appliance Library in your web browser using the following link: <https://cal.sap.com>.

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](#).

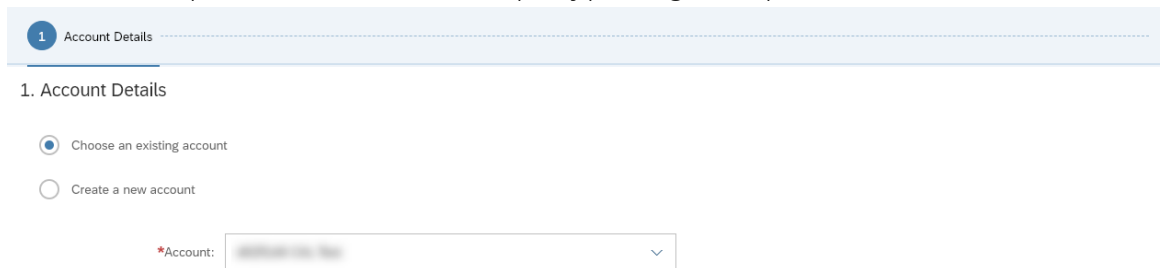
## 2.3 Creating a Solution Instance

The following steps guide you through the creation of a solution instance of SAP Data Hub, trial edition 2.3. Wherever you need more information, refer to the [documentation](#).

1. Navigate to solution SAP Data Hub, trial edition 2.3.

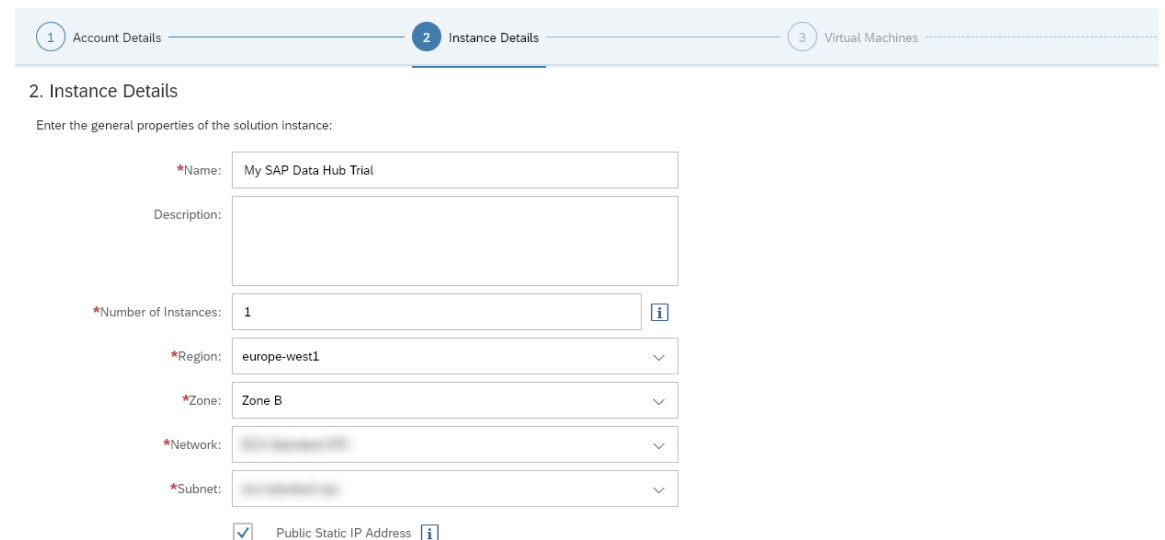


2. Press the Create Instance button and then switch to Advanced Mode (pressing the button in the lower right of the screen).
3. Select the cloud provider account. Move to step 2 by pressing the Step 2 button.



Step 2

4. Enter Name, Number of Instances (1), Region, Zone, Network and Subnet. Ensure that the checkbox Public Static IP Address is marked. Move to step 3 by pressing the Step 3 button.



Step 3



- Next check the parameters of the Jump Box and the Windows Remote Desktop (i.e. Sizes, Expandable Storage, Access Points). Normally you do not have to change anything. Move to step 4 by pressing the Step 4 button.

1 Account Details — 2 Instance Details — 3 Virtual Machines

### 3. Virtual Machines

Select size and access points of the virtual machines:

Sizes

Virtual Machine	Size	Activate
Jump Box	n1-standard-2 (2 cores, 7.50GB memory, SSD)	<input checked="" type="checkbox"/>
Windows Remote Desktop	n1-standard-4 (4 cores, 15GB memory, HDD)	<input checked="" type="checkbox"/>

Expandable Storage

Volume	Default Size	Additional Size	Total
<b>Jump Box</b>			
OS Volume	60 GB	0 GB	60 GB
<b>Windows Remote Desktop</b>			
OS Volume	50 GB	0 GB	50 GB
frontend_config	2 GB	0 GB	2 GB

Access Points Add

Virtual Machine	Service	Port Range	IP Range	Type	Enabled
Jump Box	SSH	22	0.0.0.0/0	Default	<input checked="" type="checkbox"/>
Windows Remote Desktop	RDP	3389	0.0.0.0/0	Default	<input checked="" type="checkbox"/>

Step 4

- Next check the parameters of SAP Data Hub (Kubernetes Cluster, Access Points). Normally you do not have to change anything. Move to step 5 by pressing the Step 5 button.

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

### 4. Kubernetes Cluster

\*Number of Nodes: 3

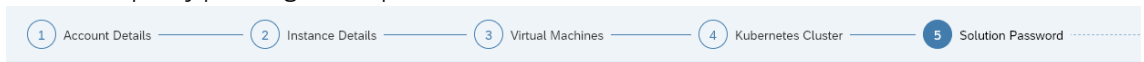
\*Size: n1-standard-8 (8 cores, 30GB memory, HDD)

Access Points

Service	Port Range	IP Range	Type	Enabled
TCP	443	0.0.0.0/0	Default	<input checked="" type="checkbox"/>

Step 5

- 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.



### 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.



### 6. Schedule Details

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

Time Zone:

#### Scheduling Options

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

Suspend Date:

#### Termination Date

- Set the termination date of the solution instance

Review

- Then (assuming you are happy with all parameters) press the Create button. It will now take around 30 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 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 Kubernetes cluster to run SAP Data Hub
- a Windows Remote Desktop from which you can easily connect to SAP Data Hub

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 explains how you can access SAP Data Hub through the Windows Remote Desktop.
- Chapter 3.4 shows how you can access SAP Data Hub without using the Windows Remote Desktop. Instead you maintain your local host file and use the web browser of your local computer.
- Chapter 3.5 outlines how to access the Jump Box for troubleshooting.
- Chapter Error! Reference source not found. finally tells you how to connect your solution instance to Google Cloud Storage.

## 3.2 Determining the external IP addresses of your solution instance

If you like to access your solution instance, you need to know the external IP addresses. You can look them up by opening your solution instance in the SAP Cloud Appliance Library. The following screenshot illustrates where to find the external IP addresses for the Jump Box, for SAP Data Hub itself and for the Windows Remote Desktop.

The screenshot shows the SAP Cloud Appliance Library interface for a 'My SAP Data Hub Trial' instance. The instance is in an 'Active' state. The 'IP Addresses' section is highlighted with a red box and contains the following information:

IP Address
SAP Data Hub External IP Address
Kubernetes Master IP Address
Jump Box Internal IP Address
Jump Box External IP Address
Windows Remote Desktop Internal IP Address
Windows Remote Desktop External IP Address

## 3.3 Accessing SAP Data Hub through the Windows Remote Desktop

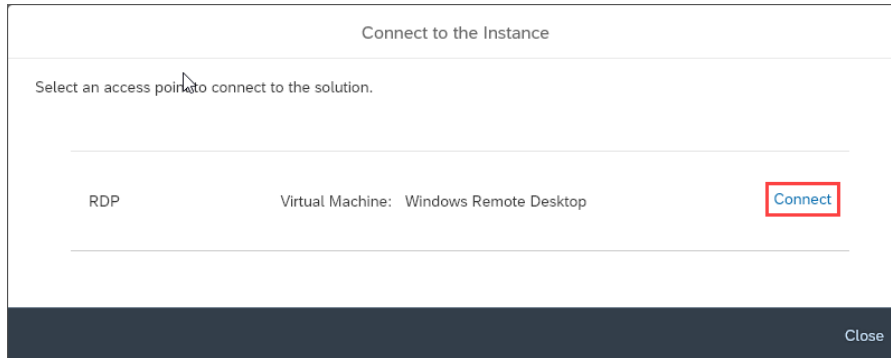
The easiest way to access SAP Data Hub, trial edition is through the Windows Remote Desktop included in the appliance. This runs in the same network and subnet as SAP Data Hub itself.

The necessary steps to access SAP Data Hub through the Windows Remote Desktop are described subsequently.

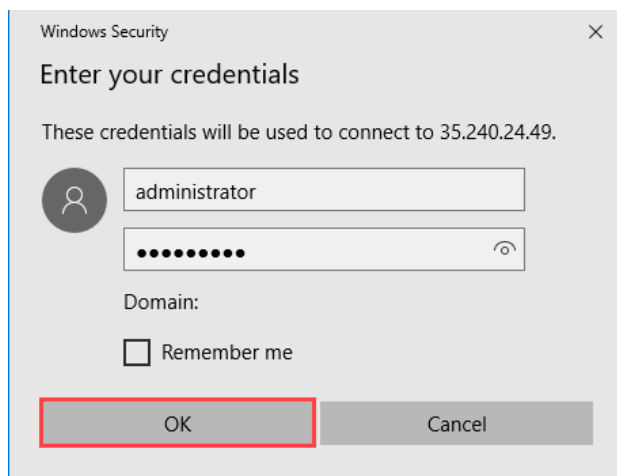
1. Press the Connect button in the SAP Cloud Appliance Library (alternatively you can look up the external IP address of the Windows Remote Desktop and initiate the remote desktop connection “manually” from your local computer).

The screenshot shows the SAP Cloud Appliance Library interface for a 'My SAP Data Hub Trial' instance. The 'Connect' button is highlighted with a red box.

- In the popup window press Connect again.

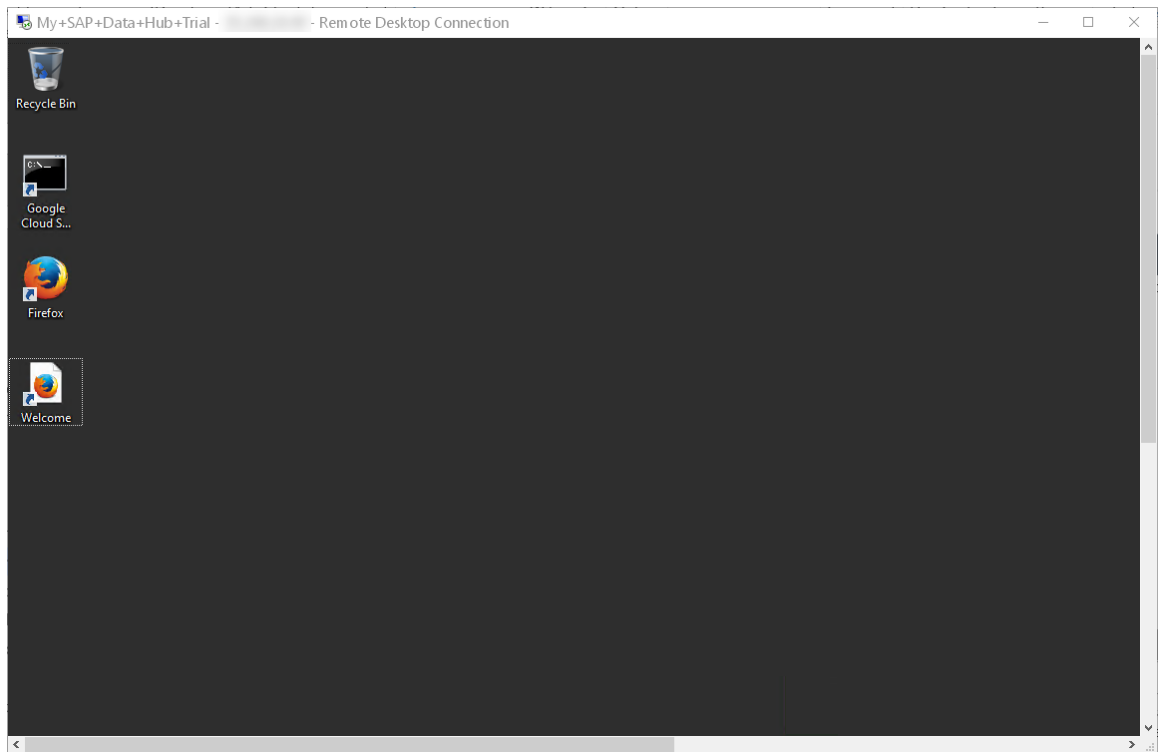


- Your web browser downloads the remote desktop connection. Open this and then logon to the Windows Remote Desktop.



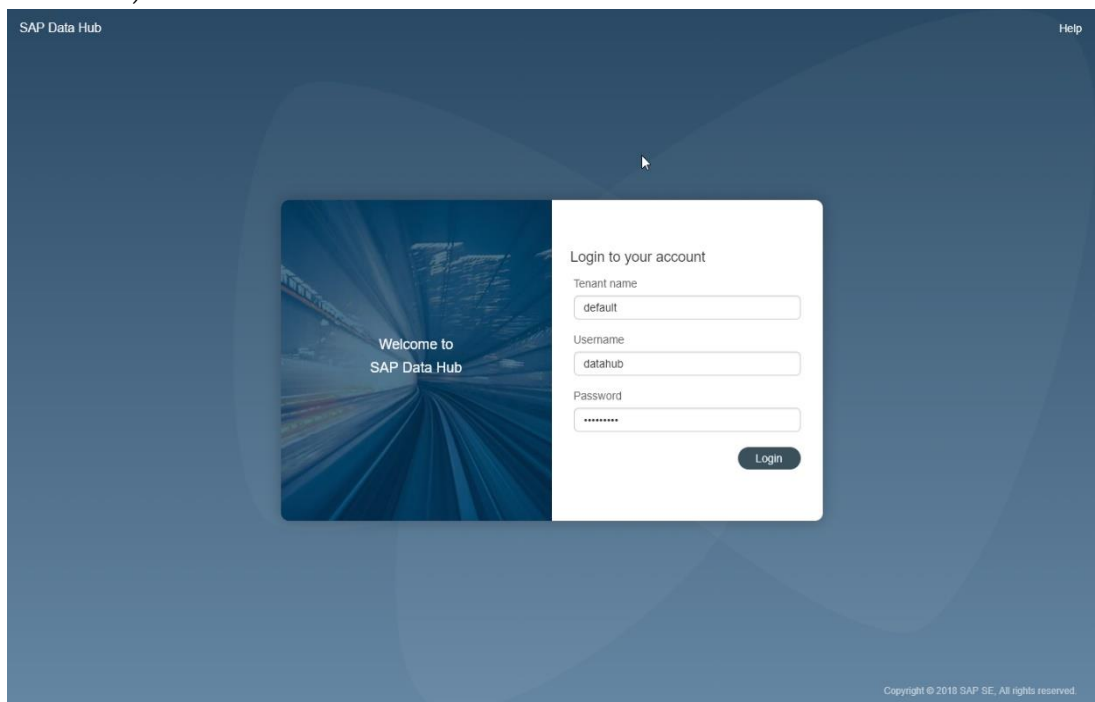
Name	Value	Description
User name	administrator	Windows Remote Desktop user
Password	<Master Password>	The initial master password of the system you provided in the SAP Cloud Appliance Library when creating the instance.

4. On the Windows Remote Desktop you find Firefox and a Welcome page. Open the Welcome page.



Attention: We recommend using Google Chrome while working with SAP Data Hub, trial edition. To download it, use this [link](#).

5. Now you can open SAP Data Hub with the web browser installed on the Windows Remote Desktop (Microsoft Internet Explorer, Google Chrome, Mozilla Firefox) by opening <https://sapdatahubtrial/>. If you see an error message that the web browser is not able to verify the security certificate of [https://sapdatahubtrial](https://sapdatahubtrial/), you need to confirm this error message (reason: the appliance uses self-signed certificates).



Name	Value	Description
URL	<a href="https://sapdatahubtrial/">https://sapdatahubtrial/</a>	SAP Data Hub logon screen
Tenant name	default system	Default tenant System tenant Attention: Typically, you use the default tenant. Use the system tenant only if you like to perform administrative tasks that require this.
Username	datahub system	SAP Data Hub user for default tenant SAP Data Hub 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 SAP Data Hub from your local computer

If you prefer to access SAP Data Hub, trial edition with the web browser of your local computer instead of using the Windows Remote Desktop included in the appliance, you need to maintain your local host file. The reason is that SAP Data Hub, trial edition uses the same hostname `sapdatahubtrial` for all solution instances. This hostname is not fully qualified and you hence need to map the actual (external) IP address against it.

The necessary steps to access SAP Data Hub from your local computer are described subsequently.

1. Look up the external IP addresses for your solution instance in the SAP Cloud Appliance Library as described in chapter 3.2 Determining the external IP addresses of your solution instance. Identify the external IP address of SAP Data Hub (SAP Data Hub External IP Address). Example: Let's assume that the SAP Data Hub External IP Address is 123.456.78.91.
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:

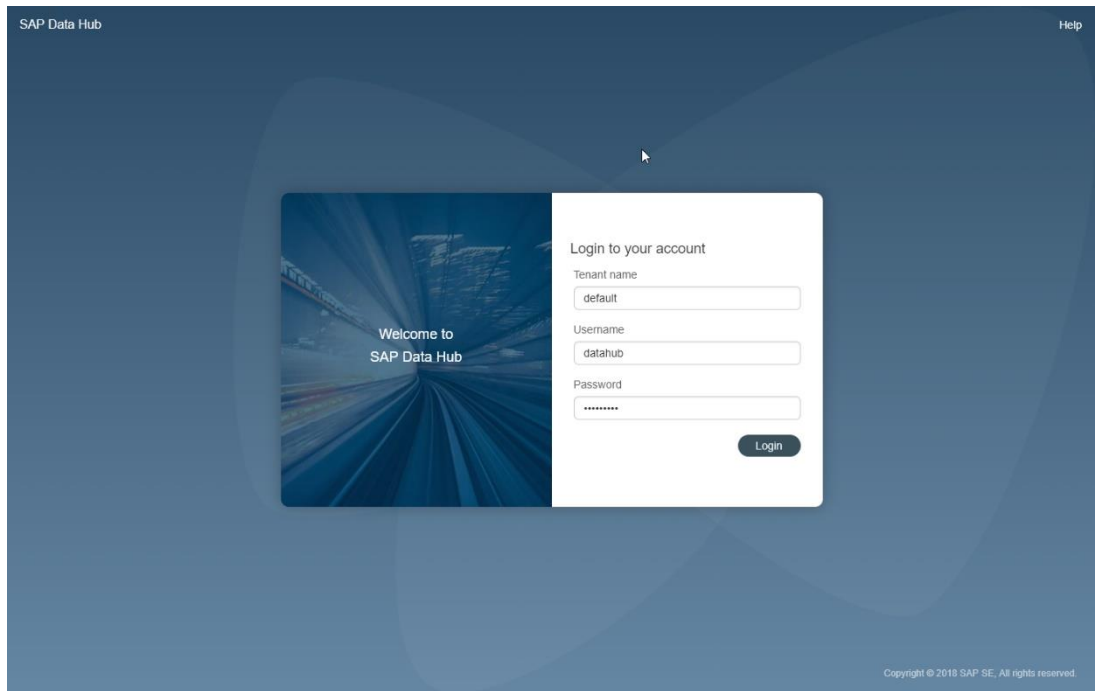
```
# SAP Data Hub, trial edition
<SAP Data Hub External IP Address> sapdatahubtrial
```

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

```
# SAP Data Hub, trial edition
123.456.78.91 sapdatahubtrial
```

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

8. Now you can open SAP Data Hub with the web browser of your local computer (Microsoft Internet Explorer, Google Chrome, Mozilla Firefox) by opening <https://sapdatahubtrial/>.  
If you see an error message that the web browser is not able to verify the security certificate of [https://sapdatahubtrial](https://sapdatahubtrial/), you need to confirm this error message (reason: the appliance uses self-signed certificates).



Name	Value	Description
URL	<a href="https://sapdatahubtrial/">https://sapdatahubtrial/</a>	SAP Data Hub logon screen
Tenant name	default system	Default tenant System tenant Attention: Typically, you use the default tenant. Use the system tenant only if you like to perform administrative tasks that require this.
Username	datahub system	SAP Data Hub user for the default tenant SAP Data Hub 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.



## 3.5 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	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 SAP Data Hub.

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-d7ae12e2-2lm1        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 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 Hub

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

---

## 6 Troubleshooting

### You cannot select a source in the “Data Transform” operator when using Mozilla Firefox

You want to use the “Data Transform” operator. You select a Connection ID. Afterwards you are not able to select a Source (the popup window stays blank).

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

Solution: Use Google Chrome while working with SAP Data Hub, trial edition (also when using the Windows Remote Desktop).

---

## 7 Appendix

### Installed Software Components

Name	Release	Support Package Stack
SAP DATA HUB - FOUNDATION	2	3

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



[www.sap.com/contactsap](http://www.sap.com/contactsap)

© 2018 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](http://www.sap.com/corporate-en/legal/copyright/index.epx) for additional trademark information and notices.