ABBYY Vantage

System Administrator's Guide

Table of Contents

About ABBYY Vantage

ABBYY Vantage is a comprehensive Content Intelligence platform that provides AI-powered cognitive services and pre-trained and trainable skills that can "understand" business documents and extract actionable data and insights.

This no-code / low-code platform makes today's digital worker and processes smarter and empowers the new citizen developer to accelerate digital transformation initiatives and expand automation to new processes in a fast and simple way, making an immediate impact on business results and customer experience.

Types of documents that can be processed with Vantage

Vantage is capable of processing structured, semi-structured, and unstructured documents in a variety of input formats and languages.

- **Structured documents** are documents which always include the exact same information in the exact same space, such as pre-defined forms, where the date has been filled out within designated areas.
- **Semi-structured documents** are documents which generally include the same or similar information, but in each document the location, size, and number of fields may vary from document to document. Examples of semi-structured documents are bills, payment orders, and invoices.
- **Unstructured documents** contain information that is not structured in any way.

The Vantage platform comes with a set of built-in Skills, which can extract data from certain document types out-of-the-box (i.e. invoices, purchase orders, receipts, bills of lading, delivery notes). These skills can be adjusted according to specific requirements and further trained based on customer-specific documents.

If the desired Document Skill is not available in the Vantage Skill Catalog more skills and technology components can be found in the ABBYY Marketplace.

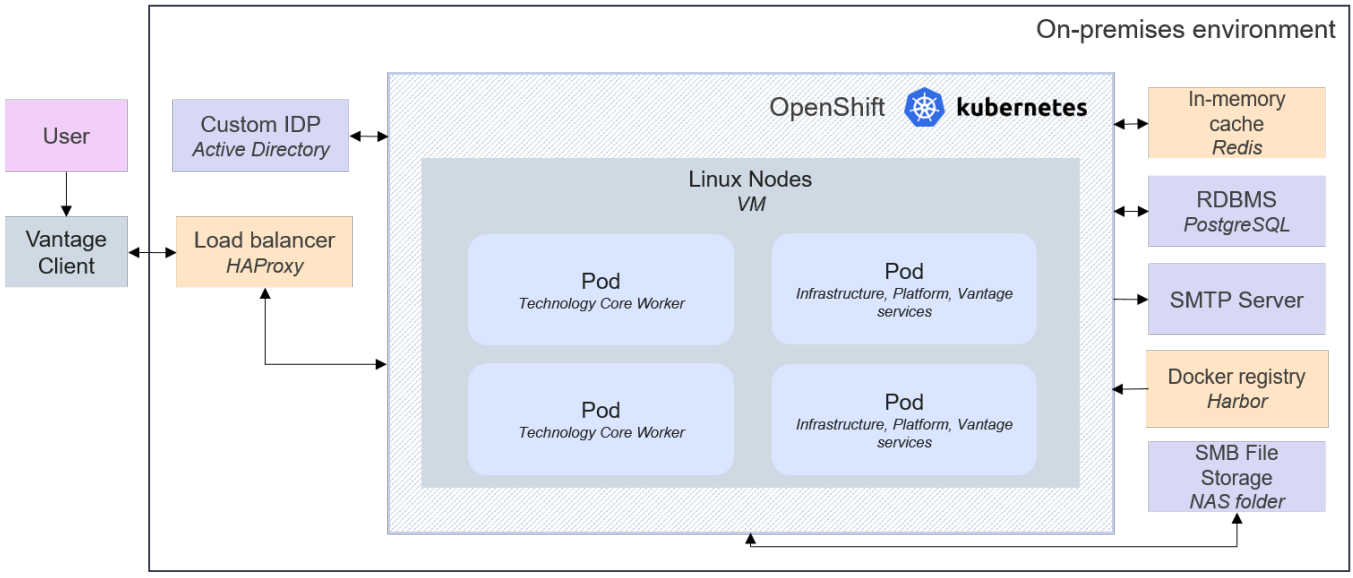
Vantage users have also the option to design and train a completely new Document Skill, Classification Skill, and/or Process Skill based on their own document set.

Installing ABBYY Vantage on an OpenShift cluster

ABBYY Vantage is provided as a set of Docker containers running in an OpenShift cluster and interacting with third-party services:

- Harbor Docker registry
- Redis in-memory cache (for Highly available configuration)
- HAProxy load balancer
- database server
- SMTP server
- file storage

The installation scheme is as follows.



ABBYY Vantage installations can be implemented in two deployment options:

- **Highly available**. A scaling configuration intended for the production process, allows you to process a large number of pages and and includes redundancy fail-safes.
- **Without high availability**. A Proof of Concept (PoC) configuration recommended for demo, test, and trial deployments, allows you to process up to 10,000 pages per 8 hours. Requirements are minimal, fault tolerance is optional.

For both deployments, to install ABBYY Vantage, you will need several virtual machines, which will become cluster nodes, and some additional services. The *Harbor*, *Redis*, and *HAProxy* services will be installed together with ABBYY Vantage. The database server, SMTP server, and file storage must be configured by the administrator in advance.

Requirements for both installation environments are listed in the System Requirements section. For instructions on how to install and configure the above services, please refer to the Installation and Initial Setup sections.

System Requirements

- OpenShift 4.8 (or later) cluster
- NAS SSD with a throughput of 1 Gbps or better, accessible via SMB
- PostgreSQL 12.2 (or later) database
- SMTP server
- TLS certificate for NGINX (either a wildcard or a domain-specific certificate issued by a **public** certification authority)
- Virtual machines:

▾ Highly available configuration

Virtual Machine	Operating System	Number	CPU	RAM, GB	Disk Storage, GB	Network
Linux nodes for OpenShift cluster	Red Hat CoreOS 4.8+	4	32	36	200	1 Gbps
In-memory cache (Redis)	Ubuntu 18.04	6	2	4	40	1 Gbps
Load balancer (HAProxy)	Ubuntu 18.04	2	2	2	40	1 Gbps
Docker registry (Harbor)	Ubuntu 18.04	1	2	4	1024	1 Gbps

▾ Without high availability configuration

Virtual Machine	Operating System	Number	CPU	RAM, GB	Disk Storage, GB	Network
Linux nodes for OpenShift cluster	Red Hat CoreOS 4.8+	4	8	16	200	1 Gbps
Load balancer (HAProxy)	Ubuntu 18.04	1	2	2	40	1 Gbps
Docker registry (Harbor)	Ubuntu 18.04	1	2	4	1024	1 Gbps

Before you begin, make sure that your administrator's machine meets the following requirements:

- Operating system: Ubuntu 18.04
- User account named "ansible" with sudo NOPASSWD:ALL privileges and access via an SSH key
- Python version 3.6 or later

Additionally:

- The private SSH key for the ansible user must be located in file ~/.ssh/ansible
- The kubeconfig file needed to connect to the OpenShift cluster must be located in file ~/.kube/config

Installation

Preparing Machines

Before you begin, make sure that your virtual machines meet the requirements listed in the System Requirements section, as well as the following additional requirements:

- Your cluster must have access to a public docker registry (docker.io, k8s.gcr.io).
- An external SQL database server must be set up and you must have the appropriate credentials to create and control new databases.
- A DNS address must be created for accessing ABBYY Vantage (e.g. vantage.yourdomain.tld).
- A DNS record must be created for accessing the docker registry (the default registry is registry.yourdomain.tld). Use the IP address specified in the **osregistry** parameter in ./ansible/inventories/os/inventory.

Preparing Dependencies

Prior to running the installer, do the following:

1. Download the installer archive and unpack it to a directory of your choice.
2. Run the following command to install the required dependencies under the "ansible" user:

```
sudo apt install -y python3-pip
sudo -H pip3 install -r requirements.txt
```

3. Specify your own values in ./ansible/inventories/os/inventory, replacing X.X.X.X with the IP addresses of the balancer virtual machines that you have created, the Linux worker nodes, the Docker registry, and the Redis cluster.

▾ High available configuration

```
[localhost]
127.0.0.1

########################################

[LoadBalancer]
oslb01 ansible_host=X.X.X.X
oslb02 ansible_host=X.X.X.X

[WorkersLinux]
osworker01 ansible_host=X.X.X.X
osworker02 ansible_host=X.X.X.X
osworker03 ansible_host=X.X.X.X
osworker04 ansible_host=X.X.X.X
```

```
[Redis]
os1redis01 ansible_host=X.X.X.X
os1redis02 ansible_host=X.X.X.X
os1redis03 ansible_host=X.X.X.X
os1redis04 ansible_host=X.X.X.X
os1redis05 ansible_host=X.X.X.X
os1redis06 ansible_host=X.X.X.X

[DockerRegistry]
osregistry ansible_host=X.X.X.X

##########################################

[linux:children]
LoadBalancer
WorkersLinux
DockerRegistry

[vm:children]
linux

[LoadBalancer:vars]
cpu=2
mem=2

[WorkersLinux:vars]
cpu=20
mem=48

[DockerRegistry:vars]
cpu=2
mem=4
```

Without high availability configuration

```
[localhost]
127.0.0.1

##########################################

[LoadBalancer]
oslb01 ansible_host=X.X.X.X

[WorkersLinux]
osworker01 ansible_host=X.X.X.X
osworker02 ansible_host=X.X.X.X
osworker03 ansible_host=X.X.X.X
osworker04 ansible_host=X.X.X.X

[DockerRegistry]
osregistry ansible_host=X.X.X.X

##########################################

[linux:children]
LoadBalancer
DockerRegistry
```

```
[vm:children]
linux

[LoadBalancer:vars]
cpu=2
mem=2

[WorkersLinux:vars]
cpu=20
mem=48

[DockerRegistry:vars]
cpu=2
mem=4
```

4. Specify your own values in /ansible/inventories/os/group_vars/all/env_specific.yml:

```
openshift_enabled: true
openshift_size: small
openshift_monitoring:
   userWorkloadEnabled: true

domain: yourdomain.tld
product_host: "vantage.{{ domain }}"

loadbalancer:
  external_ip: X.X.X.X
  virtual_router_id: 100

container_registry_host: "registry.yourdomain.tld"

platform_admin_email:  admin@yourdomain.tld

smb:
  host: X.X.X.X
  share_name: cluster_volumes
  domain: .
  user: name
  password: password
  externalHost:
    host: X.X.X.X
    share_name: import
    size: 10Gi
    storageclass: smb-sftpgo
    domain: .
    user: name
    password: password

database:
  type: postgresql
  host: X.X.X.X
  username: sa
  password: PassW0rd
```

```
smtp:
  host: X.X.X.X
  login: null
  password: ""
  port: 587
  useSSL: false
```

Parameter values

Parameter	Description
openshift_size	Specifies the deployment option. This parameter can be set to **small** (Without high availability configuration) or **medium** (Highly available configuration).
openshift_monitoring.userWorkloadEnabled	Specifies whether OpenShift monitoring should be enabled for custom projects. This parameter can be set to either **True** or **False**. If you are already using monitoring in your OpenShift cluster, then the parameter value must be **False**.
domain	The primary domain.
product_host	The DNS name that will be used to access the product. For example, if **product_host** is set to **vantage**, and your primary domain is example.com, Vantage will be accessible via the following address: vantage.example.com.
platform_admin_email	The administrator's e-mail address used to log in to the website.
loadbalancer.external_ip	An additional floating IP address for the balancer, which will be referenced by the primary domain name to access Vantage. This can be any free IP address on the subnetwork hosting the balancer virtual machines.
container_registry_host	The domain name of the Docker registry that will be created during the installation. Set to **registry.yourdomain.tld** by default.
smb.host	The IP address where the storage with SMB access is hosted.
smb.share_name	The name of the network catalog used to store Persistent Volumes data.

Parameter	Description
smb.domain	Domain hosting the user. If it is a local user, leave the value as a dot.
smb.user	The username used to access SMB storage.
smb.password	The password used to access SMB storage.
smb.externalHost.host	The IP address of the host where the SMB storage used by the folder-import service is located.
smb.externalHost.share_name	The name of the network catalog used by the folder-import service.
smb.externalHost.size	The size of the tome used for storage. The recommended size is 10 GB.
smb.externalHost.domain	Domain where the user is hosted. If it is a local user, leave the value as a dot.
smb.externalHost.user	The username used to access the SMB storage.
smb.externalHost.password	The password used to access the SMB storage.
database.type	The external SQL server type (postgresql).
database.host	The IP address of the SQL server.
database.username	The username used to connect to the database (the user must have privileges required to create databases).
database.password	The password used to access the database.
smtp.host	The IP address or the name of the SMTP server host.
smtp.login	The username used to connect to the SMTP server.
smtp.password	The password used to connect to the SMTP server.
smtp.port	The port of the SMTP server.
smtp.useSSL	Specifies whether an encrypted connection should be used. This parameter can be set to either **True** or **False**.

All other parameters are unchanged.

5. If your users are planning to import documents to Process Skills from Google and Microsoft email services, you can set up authentication via OAuth 2.0 to connect to those services using the IMAP

protocol, which will make it easier to connect to Google and Microsoft accounts as part of an Input activity in a Process Skill. To do so, get the credential data (see the Registering the Application in Google and Registering the Application in Microsoft Azure sections) and use this data to replace the null values in the ansible/files/helm/solution/charts/tokenmanagement/values.yaml file as follows:

```
OAuthClientConfigurationGoogleClientId: Google Client ID

OAuthClientConfigurationGoogleClientSecret: Google client secret

OAuthClientConfigurationMicrosoftClientId: Microsoft Azure client ID

OAuthClientConfigurationMicrosoftClientSecret: Microsoft Azure client secret
```

Note: You can also set up OAuth 2.0 authentication after installing Vantage. The correct algorithm for doing so is described in the Setting up OAuth 2.0 Authentication for Connecting to the IMAP Server section.

6. Place the wildcard certificate (along with the intermediate certificate) corresponding to the primary domain name and the key in PEM format into the following files, respectively: ./ansible/files/ssl/cert.pem, ./ansible/files/ssl/key.pem.

7. If an external authentication provider (Active Directory Federation Services) with a certificate signed by an internal certificate authority is used, place the root certificate into ./ansible/files/ssl/adfs-root.pem.

Installing Vantage

The installation of ABBYY Vantage is automatic. It is configured on individual machines using the Ansible tool, which should be installed and set up on the machine used to launch the installation. Ansible interacts with the machines using various configuration scenarios (YAML playbooks).

To install ABBYY Vantage, follow the steps below.

1. Run the following command to prepare the appropriate infrastructure (from the ./ansible/ directory under the "ansible" user):

```
ansible-playbook -i inventories/os -v playbooks/site-openshift-prepare.yml
```

The script will:

- Ask the administrator to accept the terms of the EULA.
- Carry out a preliminary check regarding the resources whose parameters are specified in the env_specific.yml file of the playbook (database, SMB share, SMTP connection).

(on virtual machines)

- Set up the load balancer (HAProxy).
- Create and set up a Redis cluster (for medium deployment configuration).
- Create and set up a Docker registry.

Note: When the installer prompts you to "Save new admin password with docker registry credentials above and enter 'yes' when done," type **yes** (without quotes). A message will be displayed containing the credentials for accessing the web interface and the registry itself.

2. Run the following command to install the product (from the ./ansible/ directory under the "ansible" username):

```
ansible-playbook -i inventories/os -v playbooks/site-openshift-deploy.yml
```

The script will:

(on the OpenShift cluster)

- Deploy the metric and log collection system.
- Deploy ABBYY Vantage.

The time required to complete the installation process will depend on the performance of the selected machines. On average, the process will take about 10 minutes. Once the installation scenario has been carried out, the services will continue downloading the required images from the external registry and deploying them on the cluster. The time required to complete this step will depend on the speed of your network and the performance of the selected machines.

Once the installation is finished:

- ABBYY Vantage will be deployed on the host machines.
- You will have a URL for the provided domain name and login and password that can be used to log in, create tenants, and process documents.

Initial Setup

After ABBYY Vantage is installed:

1. Open ABBYY Vantage and log in using the credentials of the default system administrator account displayed immediately after the installation.
2. Create a new tenant and assign a subscription (for more information, see Managing a Tenant).
3. Configure authentication with Active Directory if required (for more information, see Setting up an External Identity Provider).

Managing a Tenant

ABBYY Vantage is a multitenant system that operates using the following two scopes: **System Administrator's Scope** and **Tenant's Scope**.

Note: To ensure the different clients are logically isolated, users with permissions and roles for one of the scopes are not provided access to the other, and vice versa. Similarly, tenants are not able to access data from other tenants.

The **System Administrator's Scope** covers system administration. Only users with the **System Administrator** role are able to manage the system tenants.

This section contains instructions on how to manage a tenant:

Creating and Deleting a Tenant

Creating a tenant

To create a new tenant, do the following:

1. Navigate to the **Tenants** tab in the left pane.
2. Click **New Tenant**.
3. In the dialog that will open, specify the following for your tenant: name, description, and email address of the tenant administrator. Next, upload the subscription you were provided with by clicking **Upload Subscription File**. For more information about subscriptions, see Subscriptions.

Once you have completed the above steps, a new tenant will be created in the system, and an invite with a registration link will be sent to the specified email.

A list of all tenants created in the system can be viewed in the **Tenants** tab. Tenants with invites that have not yet been accepted by the tenant administrator are denoted with an icon in front of their names. Hovering the mouse cursor over the icon displays the date when the invite was sent.

Note: Invites remain valid for 14 days, during which the user has to register by clicking the link in the invite. Otherwise, the tenant will be deleted from both the system and the list, and the link will become invalid.

If required, you can also send the same invitation again by doing the following:

1. Select the appropriate tenant by marking it in the tenant list.
2. Click **Resend Invite**.
3. In the dialog box that will open, modify the email address if required, and click **Send**.

Deleting a tenant

To delete a tenant, do the following:

1. Select the appropriate tenant by marking it in the tenant list.
2. Click **Delete** and confirm the deletion.

Once you have completed the above steps, the tenant will be removed from the list, and access to its data will be blocked. According to the ABBYY retention policy, tenants are deleted from the system 30 days after their deletion from the list. Before that period has elapsed, it is not possible to create a new tenant with the same name.

Subscriptions

For a tenant to work with ABBYY Vantage, it needs to have an active subscription. Tenant subscriptions are used to determine which application features the tenant will have access to.

In particular, tenant subscriptions determine the following:

- the amount of time during which access to skills will be provided,
- the number of pages that can be processed while the subscription is active,
- the available skills.

To count pages processed using various skills, counters are included in a subscription. For more information, see How Pages Are Counted in ABBYY Vantage.

Note: A subscription applies to all users in a tenant. If an ABBYY Vantage subscription is tied to several tenants within a single Vantage installation, these tenants will have shared counters for each skill used as part of the subscription.

Subscription Parameters

A system administrator can view information about any tenant's subscription.

This can be done as follows:

1. Click the name of the appropriate tenant in the **Tenants** tab.
2. In the dialog that will open, navigate to the **Subscription** tab.

This tab contains an overview of the following subscription parameters:

- the subscription type, its serial number and expiry date;
- counters included in the subscription;
- information about remaining available pages for each of the skills;
- the date of the next counter update if counters are renewable;
- the number of pages that will become available with the next update;
- any additional subscription options.

How Pages Are Counted in ABBYY Vantage

ABBYY Vantage keeps count of all document pages processed by users with license-specific counters. The counter type used to calculate the number of processed pages depends on the skill used and on whether the skill is licensed or not. Licensed skills are skills that have been created by ABBYY or its partners and have undergone the licensing procedure at ABBYY.
There are several types of counters used in ABBYY Vantage:

1. **Licensed Skill** counters are used to count pages processed using licensed skills. Each licensed skill has its own special counter.

2. A **Core Cognitive Skills** counter is used to calculate the number of pages processed by all unlicensed ABBYY Vantage skills.
3. An **OCR** counter is used to calculate the number of pages processed using OCR skills.
4. **Licensed Document and Classification Skills** counters are used to calculate the number of pages processed by trial versions of licensed skills only. This lets you process documents using licensed skills whose counters are either not part of your license or have run out of pages to process. This way, you can try out a licensed skill before purchasing it.

All pages processed by non-licensed skills are counted using a single Core Cognitive Skills counter. Pages processed by licensed skills are counted using separate Licensed Skill counters.

Pages are counted for documents processed during design time when modifying document and classification skills and during runtime when using any type of skill (document, classification, process, OCR).

Design time document processing for document and classification skills includes the following:

- importing demo documents to a document skill,
- importing files to a document set for training document and classification skills.

Runtime document processing includes the following:

- uploading pages to the **Documents** section,
- processing documents in transactions,
- document uploads when using the **Try Skill** feature.

All document pages processed during both design time and runtime are counted in the corresponding skill counters.

All subscription counters are updated once during a set period of time called the renewal period. At the end of this period, all the counter values are reset and the maximum number of pages again becomes available to the user.

The duration of the renewal period and its start date are the same for all counters in a subscription, regardless of when any of the skills were used for the first time.

A single image of a page of any format is counted as one page, even if the image contains several different documents.

If the number of pages available for processing by a specific skill has run out, or if the currently active license does not provide access to the required processing skill, users will not be able to create new transactions using such skills.

How pages are counted for process skills

Generally, process skills contain one or more OCR, document and/or classification skills (which can be both licensed and unlicensed). Depending on the exact set of skills contained within a process skill, pages for the document being processed will be counted as follows:

- **If the process skill uses only unlicensed skills**
  Each page will increment the Core Cognitive Skills counter once, regardless of the number of skills used to process it.
- **If the process skill uses one licensed skill**
  Each page will increment the licensed skill's counter once, regardless of how many times the licensed skill was used to process the document.
- **If the process skill uses unlicensed skills and one licensed skill**
  Each page will increment the licensed skill's counter. These pages will not increment the Core Cognitive Skills counter.
- **If the process skill uses several licensed skills**
  Each document page will increment each licensed skill's counter.

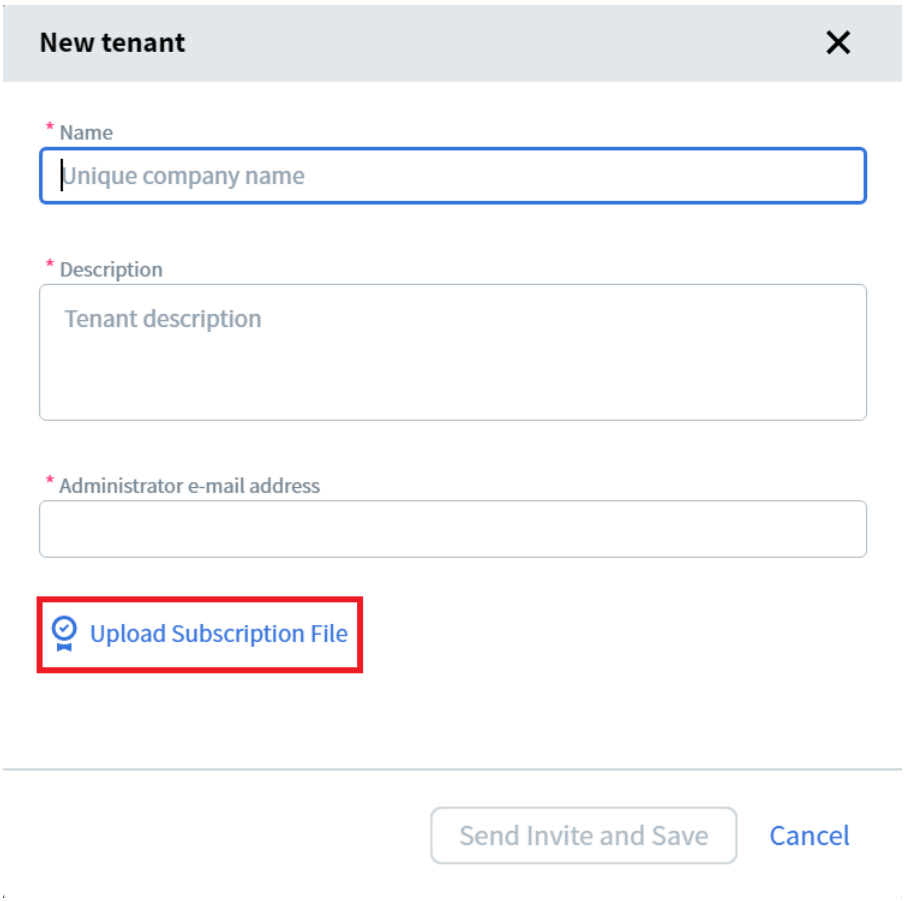
Managing Subscriptions

The system administrator can manage tenant subscriptions as follows: link, modify, and delete subscriptions.

Linking a subscription

A subscription file that you have received from ABBYY can not only be linked to a new tenant when it is created, but also to an existing tenant.

In the case of the former, the subscription file needs to be uploaded when a new tenant is created. To do so, click **Upload Subscription File**.



Modifying or deleting a subscription

If the number of pages available for processing for each of the skills has run out, or if the currently active tenant subscription lacks the required skill to process documents, the system administrator may swap the current subscription file for a new subscription file with different page limits and counters. If the tenant needs to be suspended, the subscription can be deleted.In this case, tenant users will not be able to create transactions using any of the skills.

To modify or delete a subscription, do the following:

1. Open the **Tenants** tab and click the name of the appropriate tenant.
2. In the dialog that will open, navigate to the **Subscription** tab.
3. Click the ⋮ icon displayed next to the subscription type name and select either **Change Subscription File** or **Delete Subscription**, depending on your desired action.

Setting up an External Identity Provider

ABBYY Vantage supports single sign-on authentication via the Active Directory and Azure Active Directory Identity Providers.

The authentication procedure consists of the following:

- A user navigates to the ABBYY Vantage page and enters their login credentials. Vantage then looks for suitable tenant to authenticate the user. If there are several such tenants, the user will be explicitly asked to select one.
- If an external Identity Provider has been set up for a tenant, users in that tenant will be redirected to the login page of the Identity Provider (Active Directory or Azure Active Directory). Otherwise, the user will be asked to enter their password and authenticate using Vantage Identity Provider.
- The user is authenticated via the external Identity Provider, after which they are redirected back to Vantage.

This section describes the correct procedure for setting up an ABBYY Vantage tenant, as well as how to set up the Active Directory or the Azure Active Directory service to be used as an External Identity Provider. You will need to do the following:

1. Prepare the external Identity Provider (Setting up Active Directory or Setting up Azure Active Directory).
2. Connect the external Identity Provider to the Vantage tenant (Setting up a Tenant).

Note: This setup can only be carried out by a user with the System Administrator role.

Setting up Active Directory

Prerequisites

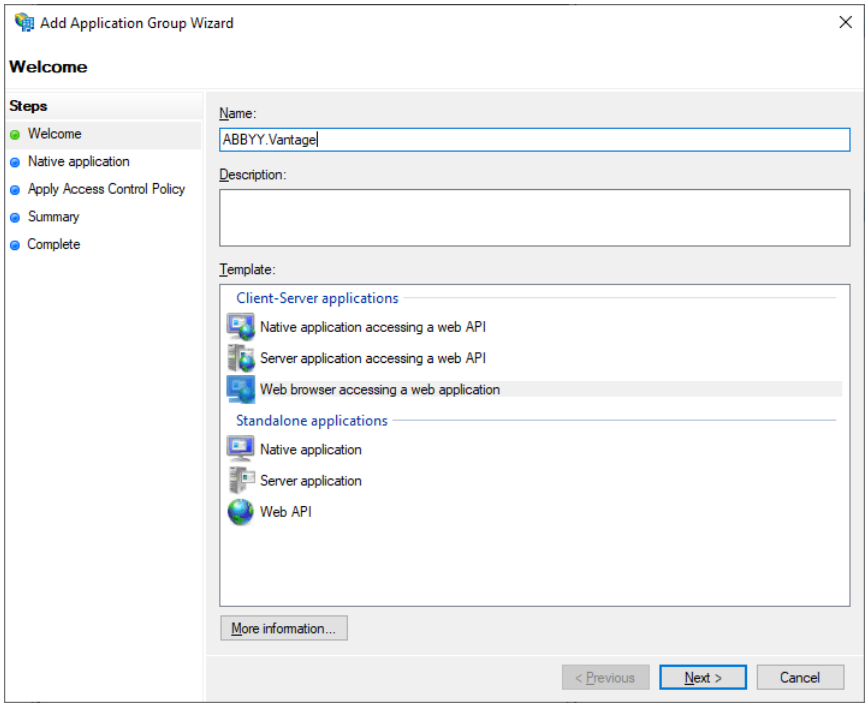
- The Active Directory Federation Service (ADFS) needs to be installed.
- A user group should be created in Active Directory. This group will be used to manage the list of users permitted to access Vantage.

You will also need a Redirect URI to receive the authentication responses. By default, the format of the URI is as follows: <VantageUri>/auth2/signin-oidc. For example, if Vantage is accessed via https://vantage-us.abbyy.com/auth2, the URI will be as follows: https://vantage-us.abbyy.com/auth2/signin-oidc.

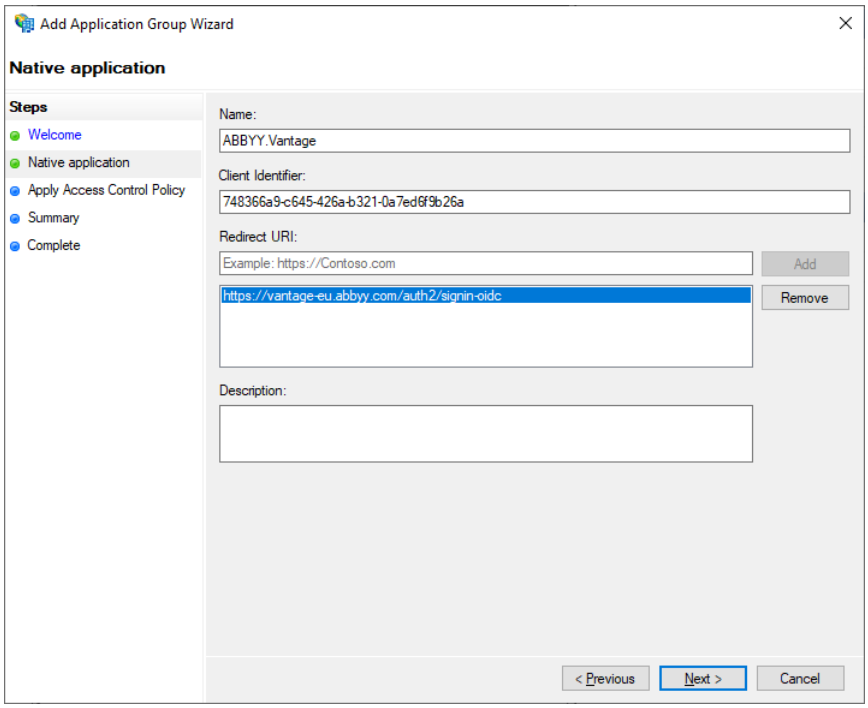
Setup

To set up Active Directory via the ADFS management console, do the following:

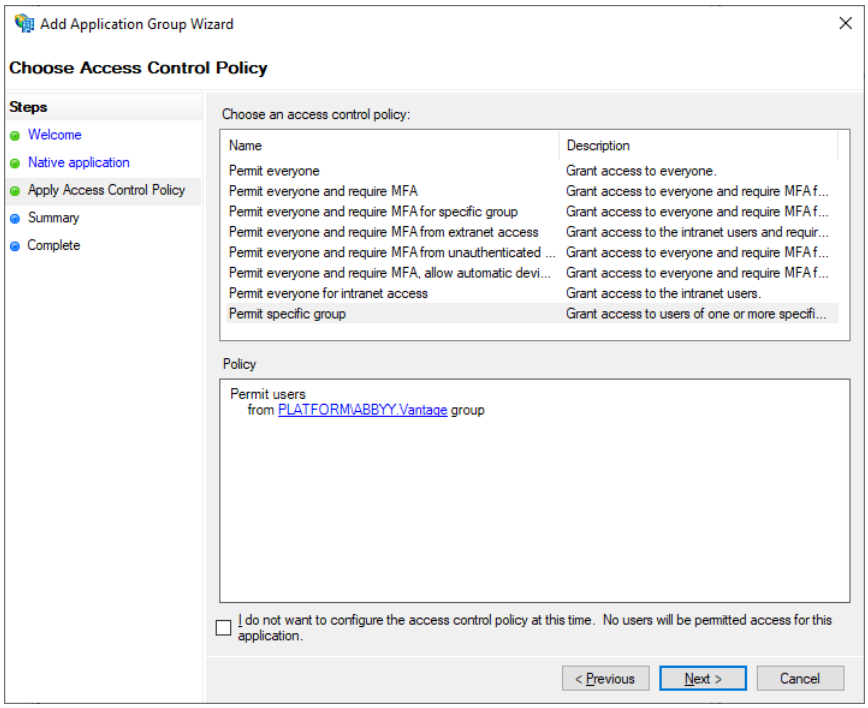
1. Open the management console.
2. Add a new application group and select the **Web browser accessing a web application** template.



3. In the **Native Application** tab, enter your Redirect URI and save the generated Client Identifier, which will be required later on. You can also view it again later in the application group properties.



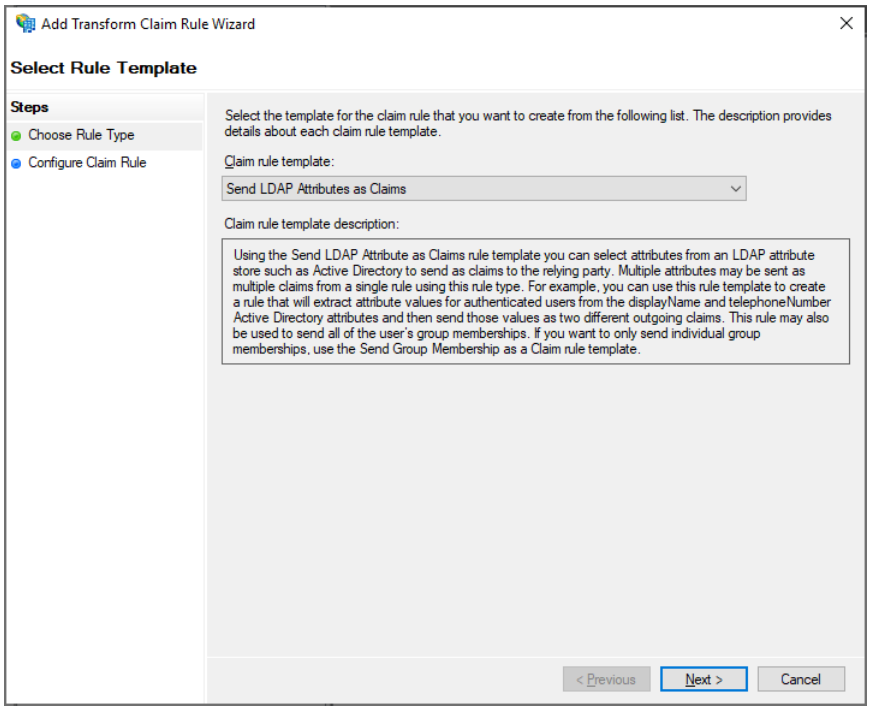
4. In the **Apply Access Control Policy** tab, select a user group to be granted access to Vantage. You can leave the default value of **Allow everyone** if you do not wish to restrict access for users at that moment.



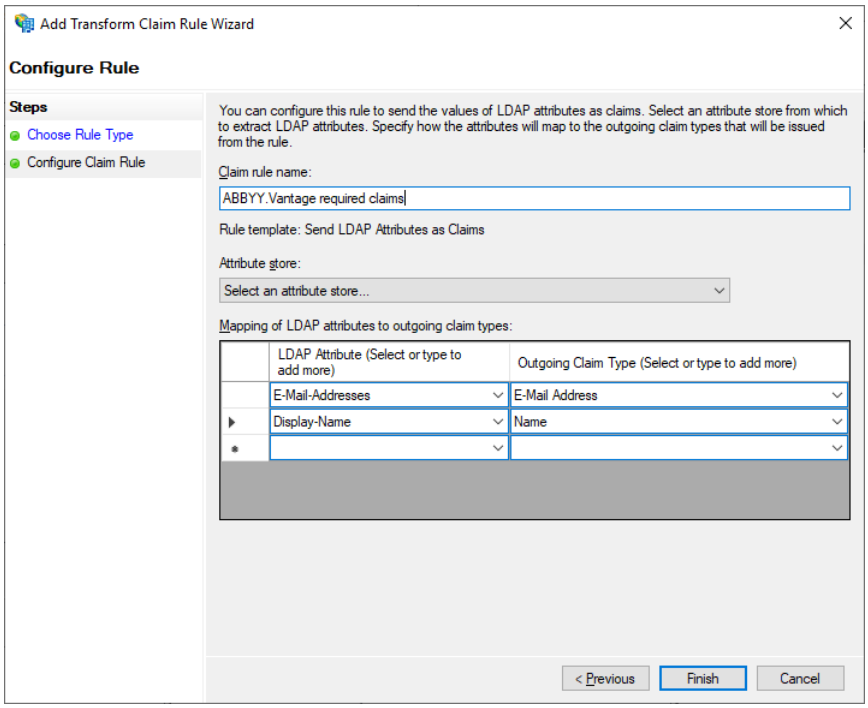
5. The **Summary** and **Complete** tabs are not required to set up Active Directory. Navigate through them and click **Close**.

6. Check the properties of the application group you have created. You can also manage Redirect URI and Client ID via **Server application Properties**. For advanced application group settings, use **Web application Properties**. There, you can also find the Application ID and save it, since it will be required later on in the setup.

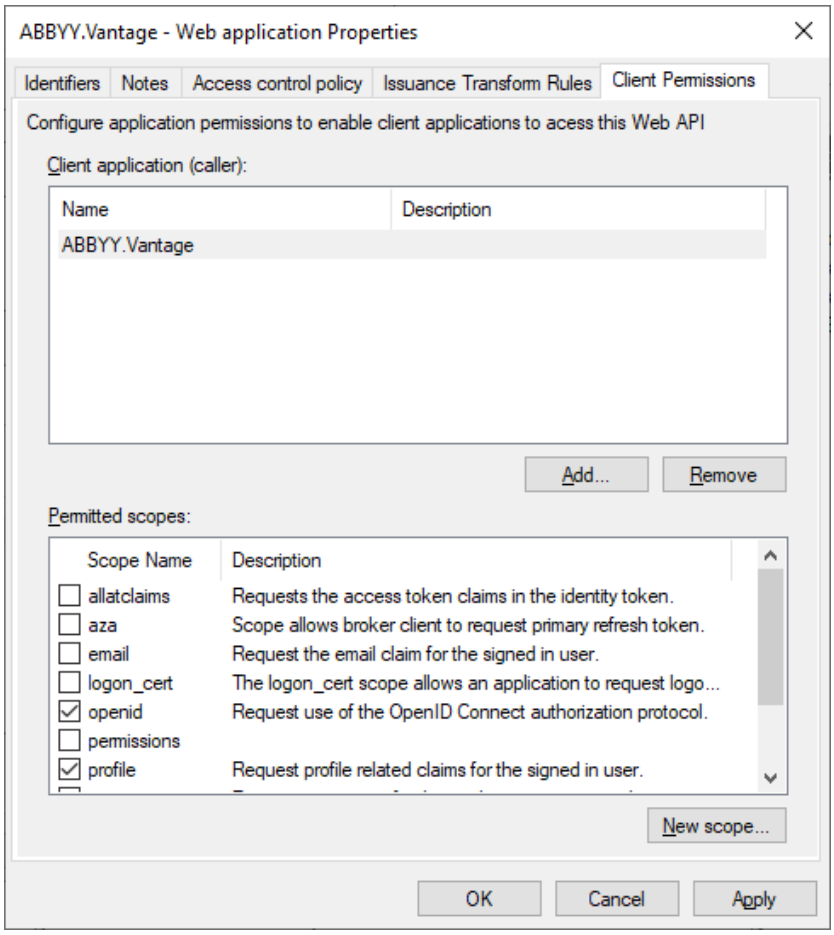
7. In the **Add Transform Claim Rule** dialog, add a claim rule to make sure that the e-mail address and name will be included in the token.



8. In the **Configure Claim Rule** tab, select Active Directory in **Attribute Store**.



9. In the **Web application Properties** dialog, navigate to the **Client Permissions** tab, select the **openid** and **profile** scopes, and click **Apply**.



This concludes the Active Directory setup. After you have completed the above steps, you will need to set up authentication via an external Identity Provider for your tenant in Vantage, which will require the following:

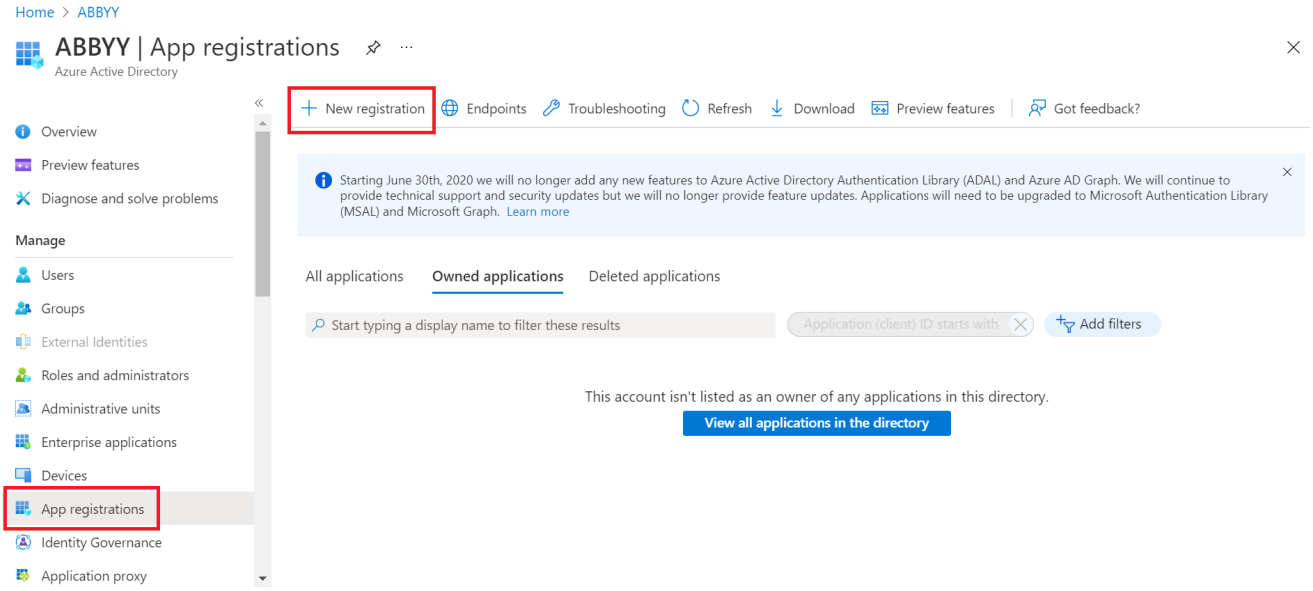
- The **Application (Client) Id** from step 6.
- The ADFS URL in the following format: https://<Full computer name>/adfs (a machine's **Full computer name** can be found in its system settings). For example, https://adfs.platform.local/adfs.

For more information, see Setting up a Tenant.

Setting up Azure Active Directory

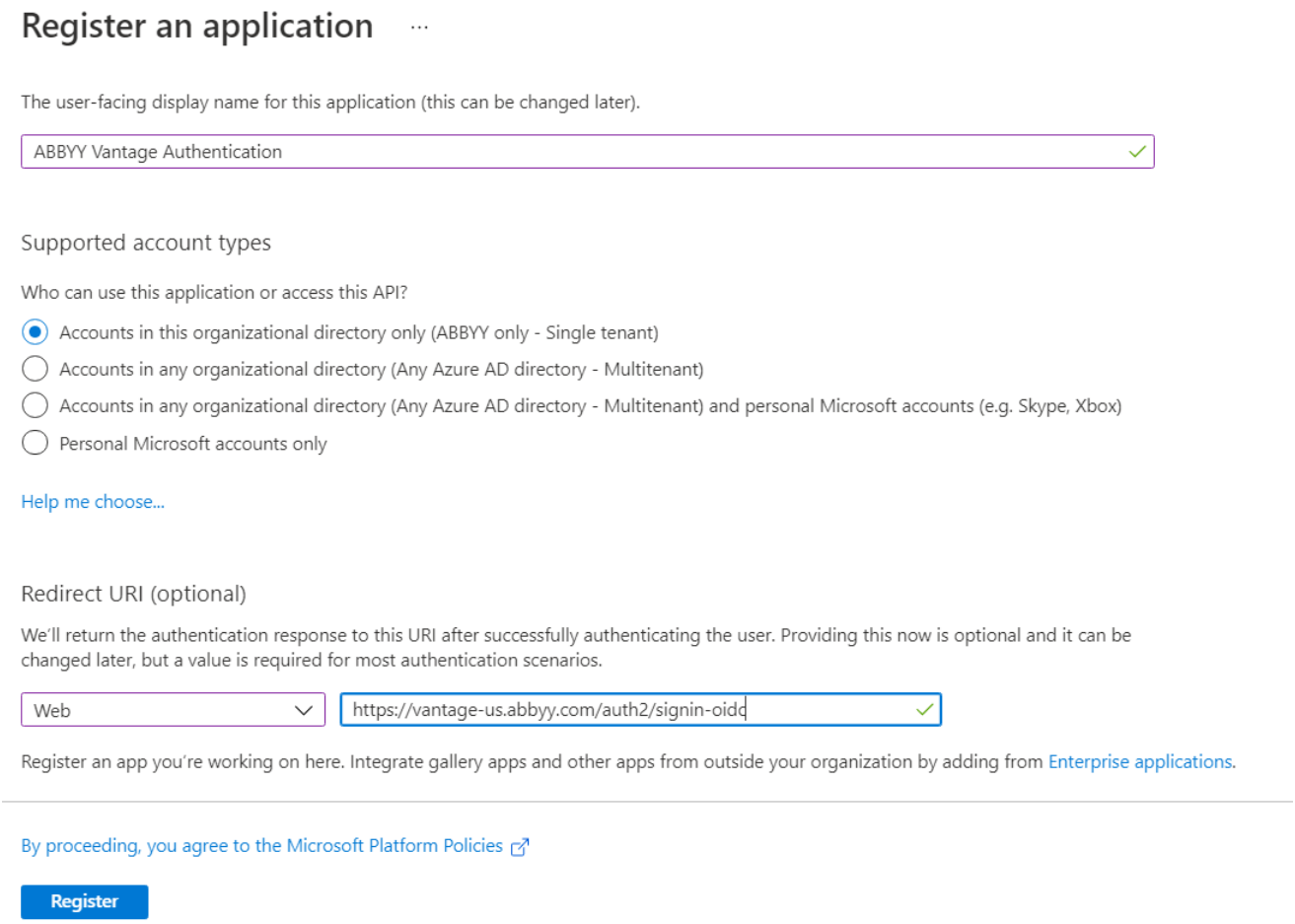
To set up Azure Active Directory, do the following:

1. Open Azure Portal and undergo authentication. In the pane on the left, select **Azure Active Directory**.
2. In the pane on the right, select **App registrations** and click **New registration**.



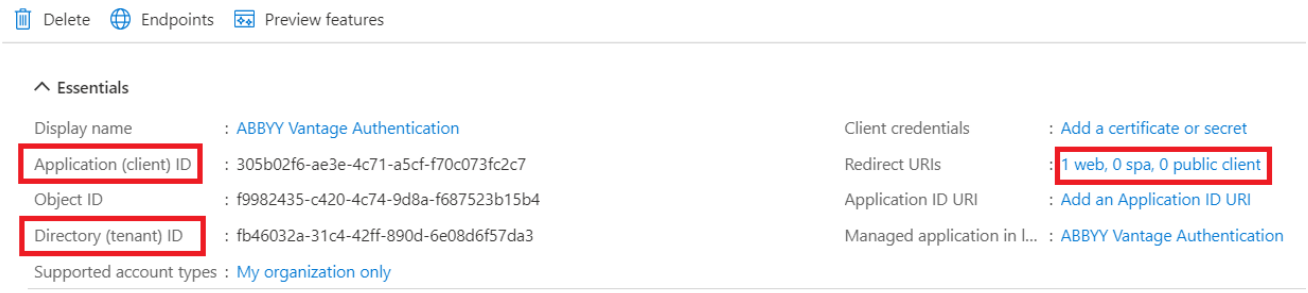
3. Fill in the required fields:

   - specify a name;
   - select **Accounts in this organizational directory only (ABBYY only - Single tenant)**;
   - specify a URI for each Vantage URL that should be able to authenticate using this account: <VantageUri>/auth2/signin-oidc.

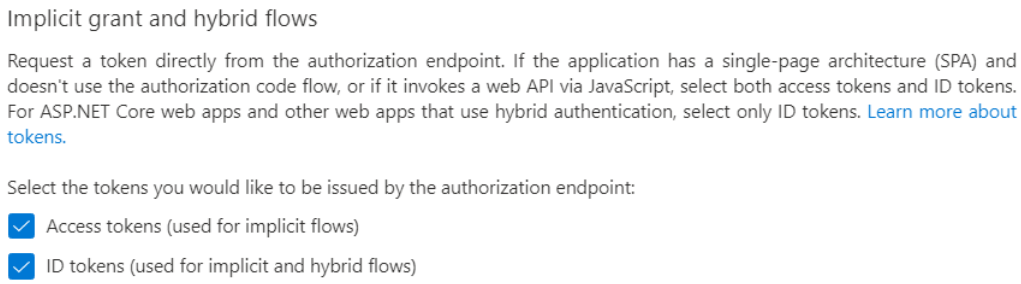


Click **Register**.

4. Save the values of Application (client) ID and Directory (tenant) ID, and then click the value of Redirect URIs.



5. Select **Access tokens** and **ID tokens**. Click **Save**.



This concludes the Azure Active Directory setup. After you have completed the above steps, you will need to set up authentication via an external Identity Provider for your tenant in Vantage, which will require the following:

- The **Application (Client) Id** from step 4.
- The ADFS URL in the following format: https://login.microsoftonline.com/<**Directory (tenant) ID**>.

For more information, see Setting up a Tenant.

Setting up a Tenant

To set up an External Identity Provider for a tenant in Vantage, do the following:

1. Use your ADFS url/Azure AD url and Application (client) ID to send the following PUT request:

```
PUT https://<product_host>.<domain>/api/publicapi/v1/tenants/{tenantId}/external-providers

{
"providerSettings": {
    "kind": "AzureActiveDirectoryOidc",
    "settings": {
        "Authority": "<ADFS url/Azure AD url>",
        "ClientId": "<Application (client) ID>"
    }
  }
}
```

2. Set up an email address domain for the tenant users. This domain should be a part of your users' email addresses, e.g. "abbyy.com". When logging in using an email address from the specified domain, users will be asked to authenticate for that tenant. Send the following PUT request:

```
PUT https://<product_host>.<domain>/api/v1/tenants/{tenantId}/custom-domains

{
  "domains": [
    "<domain part of your users emails>"
  ],
  "ownerId": "{tenantId}"
}
```

Note: The email address domain is unique for all tenants: the same email address domain cannot be used by more than one tenant to authenticate users via an external Identity Provider. However, several domains can be specified for a single tenant.

Setting up a Database Connection

ABBYY Vantage uses databases hosted on external servers and may become inoperable if those servers fail. In this case, the system administrator is able to restore such databases on a different server and set up a connection to the new databases using Consul.

Note: Before starting, make sure that the kubectl command line tool is installed and that a connection to the Kubernetes cluster has been established.

To set up a connection to a new database in the ABBYY Vantage settings, do the following:

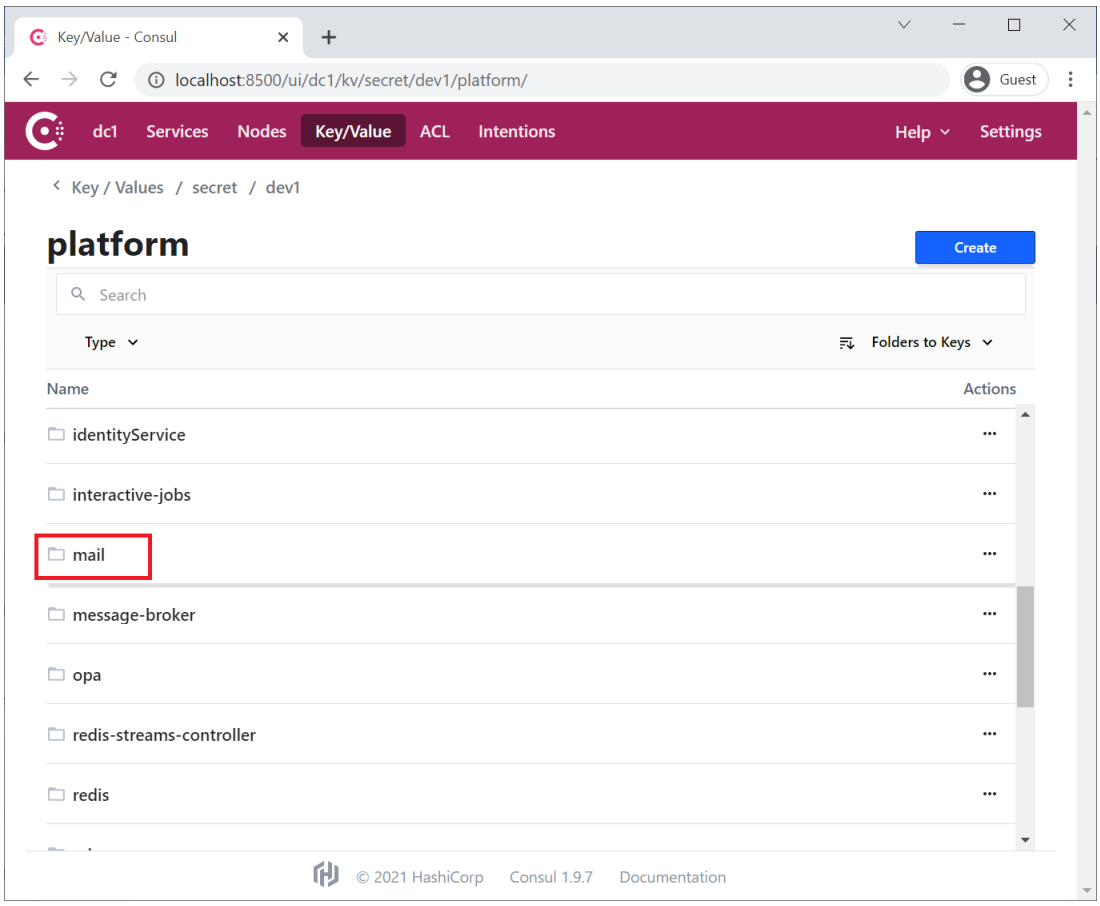
1. Access the Consul web interface by running the command below

   ```
   kubectl port-forward -n infrastructure service/consul-ui 8500:80
   ```

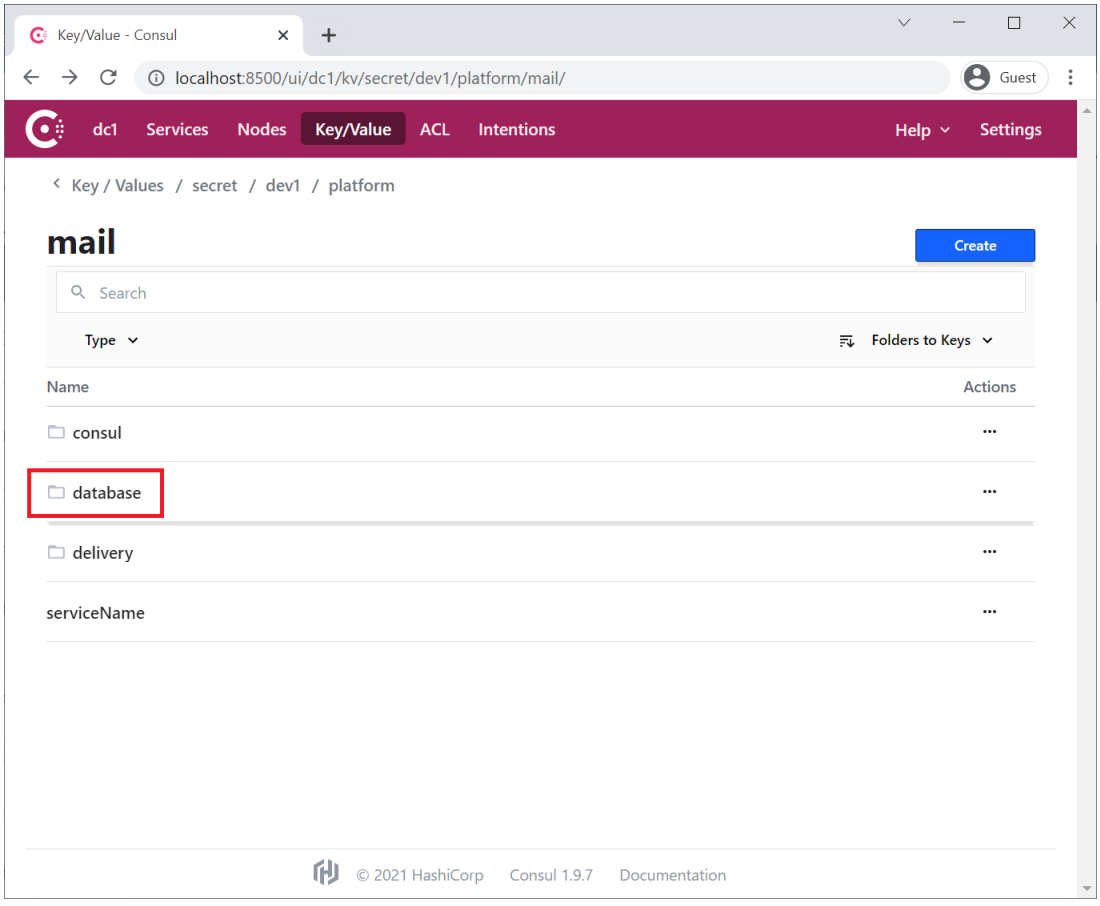
   and then navigating to http://localhost:8500/ui/dc1/kv/secret/.

2. Use the **Key/Value** tab that will open to select the correct Vantage environment.

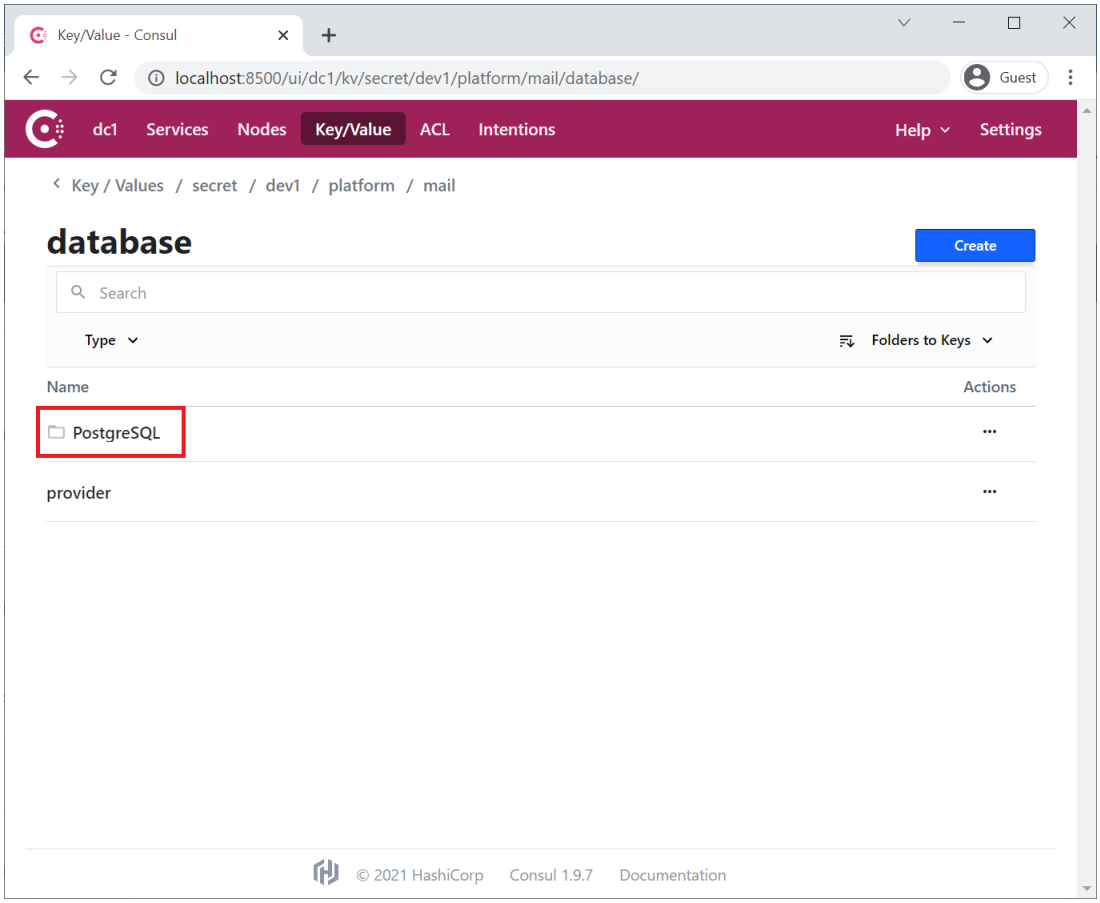
3. Select either the **platform** or the **vantage** project, as well as the appropriate service that uses the database, e.g. **mail**.



4. Navigate to the **database** section that every service contains.

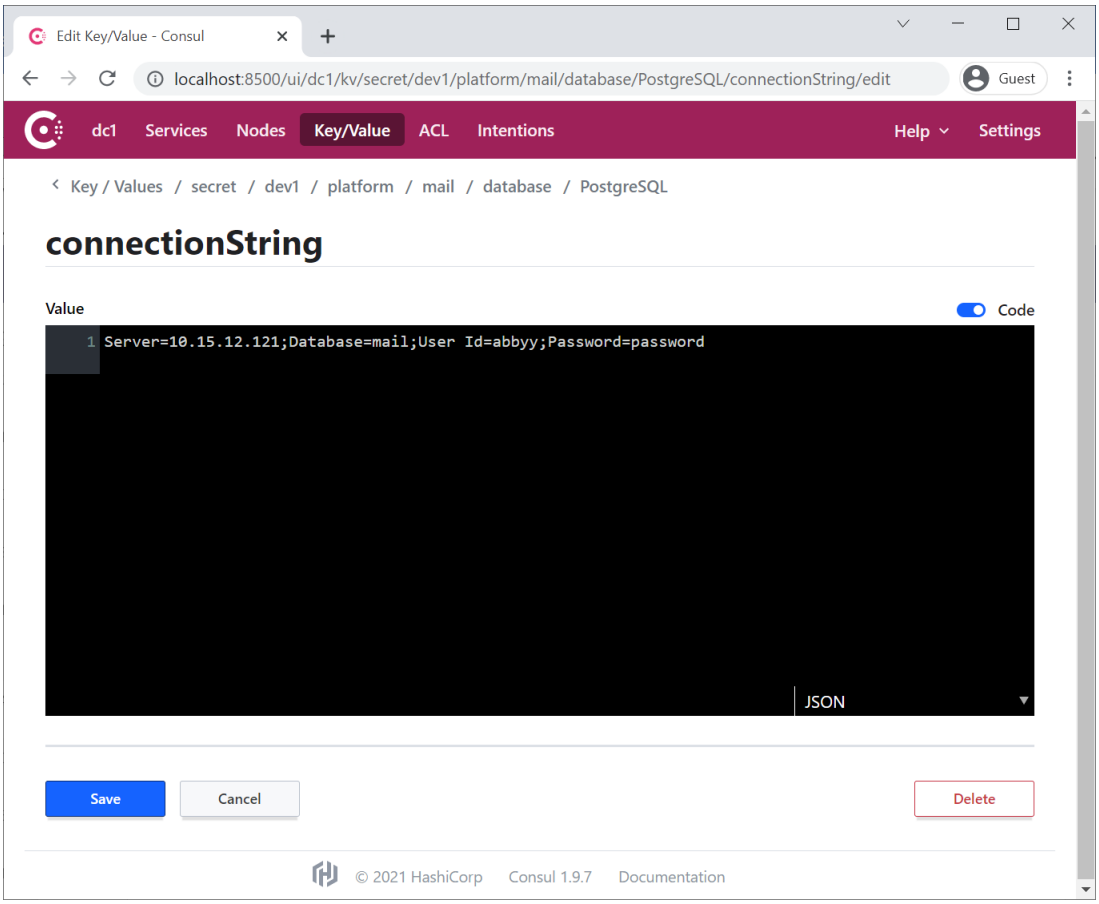


5. Open the **PostgreSQL** section.



6. In the **connectionString** key:

    1. Replace the old value of **Server** with the address of the new server.
    2. Specify the new database in the **Database** parameter.
    3. Specify the login credentials for the database in the **User Id** and **Password** parameters.



7. Click **Save**.
8. Restart the modified service by running the following command:

```
namespace=dev
label=mail
kubectl -n $namespace rollout restart $(kubectl -n $namespace get deployments -l
app.kubernetes.io/component=$label -o name)
```

Note: When a server address changes, this procedure has to be carried out for every database.

Below is a table listing all services that use the database, as well as their label that can be used to find and restart each service.

Name of the Consul section name	Service label	Notes
platform		
api-gateway-registry	api-gateway-registry	
api-registry	api-registry	
auth-adminapi2	auth-adminapi2	
auth-identity	auth-identity	

Name of the Consul section name	Service label	Notes
auth	auth-sts-identity, auth-adminapi2	This database is used by two services.
blob-storage	blob-storage	
cron-service	cron-service	
documentsetstorage	documentsetstorage	
mail	mail	
skill-monitor	skill-monitor	
storage	storage	The **database** section is stored in the **fileMetadata** catalog.
workflow-facade	workflow-facade	
vantage		
catalogstorage	catalogstorage	
folderimport	folderimport	
mailimport	mailimport	
onlinemlservice	onlinemlservice	
publicapi	publicapi	
secretstorage	secretstorage	
skill-monitor	skill-monitor	
skillinfo	skillinfo	
subscriptions	subscriptions	
tokenmanagement	tokenmanagement	
transactions	transactions	
workspace	workspace	

Setting up OAuth 2.0 Authentication for Connecting to the IMAP Server

By default, users setting up document import from an email service using an Input activity in a Process Skill only have access to basic IMAP server authentication. In order for Google and Microsoft email services authentication to also become available via the OAuth 2.0 protocol, do the following:

1. Register the applications on Google Cloud Platform and the Azure portal;
2. Generate account credentials for these applications (Client ID and Client secret);
3. Pass the generated credentials to Consul.

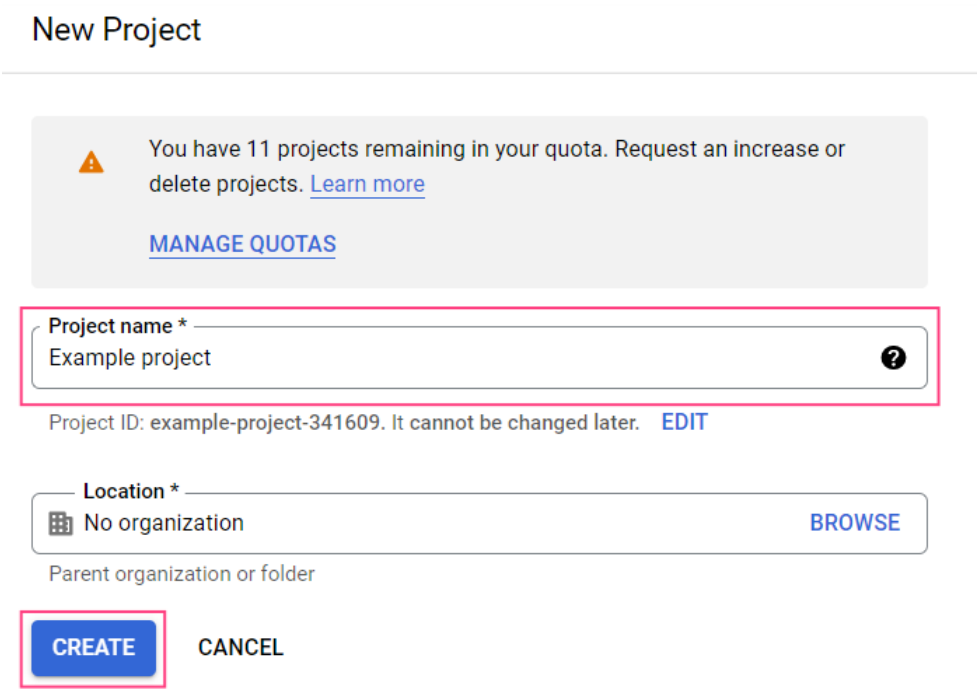
The above can be done both when preparing to install Vantage, as well as after the installation.

Registering the Application in Google

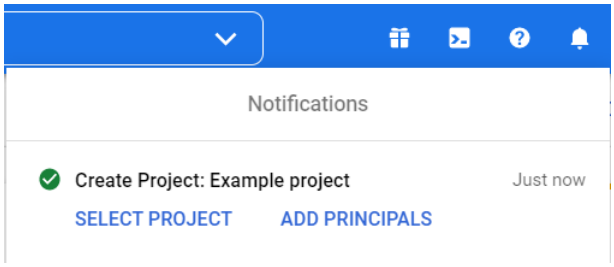
Creating an application requires a Google account.

Creating a project on the Google Cloud Platform

1. Navigate to the Google Cloud Platform New Project page.
2. Specify a name for your project and click **Create**.

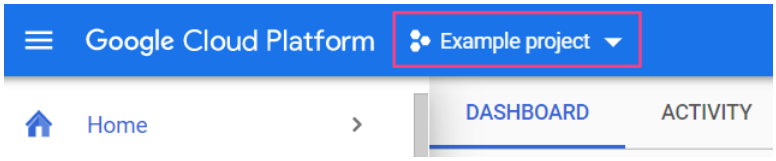


Wait for a notification saying that your project has been created.

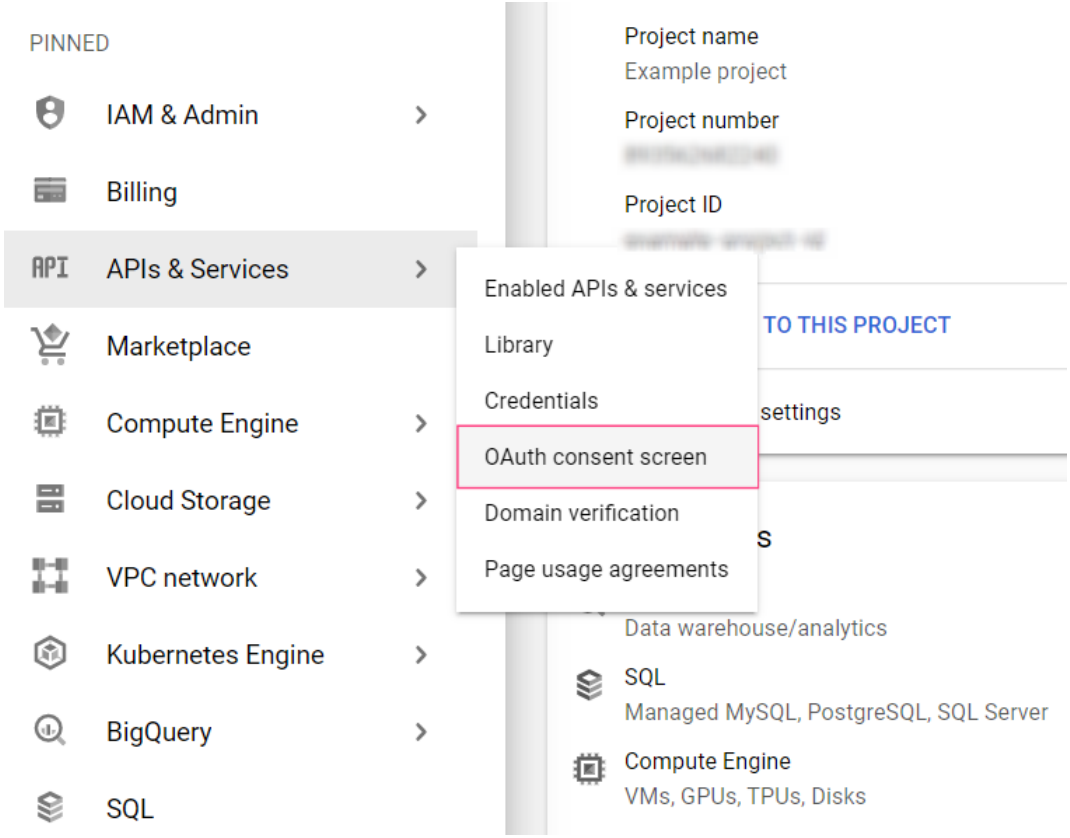


Setting up the application

1. Navigate to the Google Cloud Console and select the appropriate project.

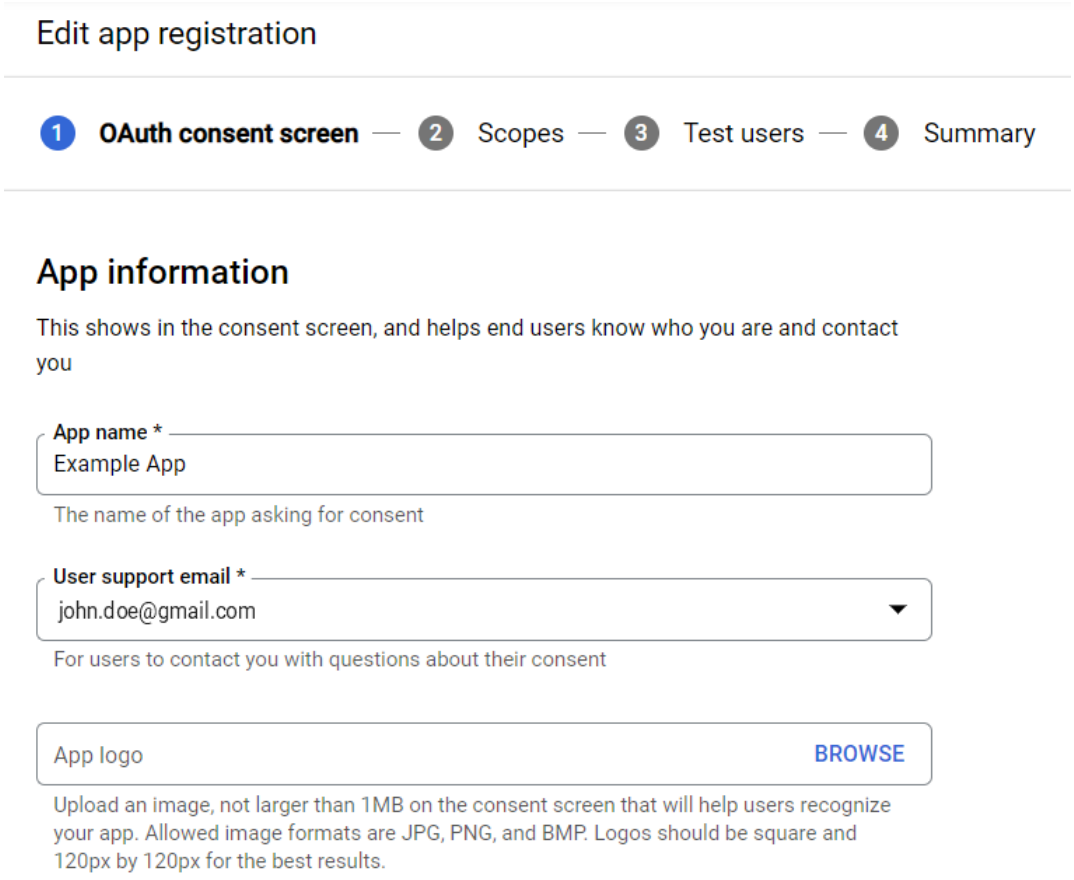


2. In the menu on the left side of the screen, select **APIs & Services** > **OAuth consent screen**.

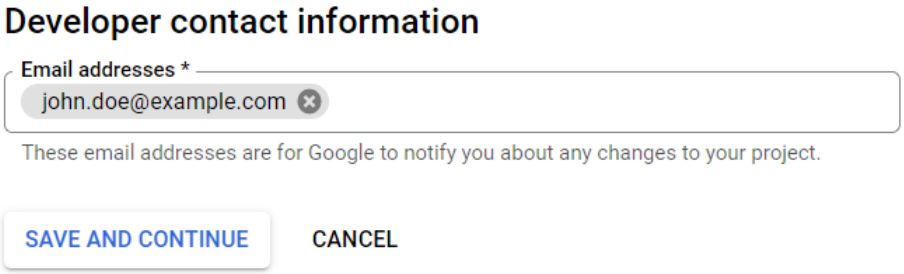


3. Select the **External** user type and click **Create**.

4. Specify a name for your application. In the **User support email** drop-down list field, select your Gmail address.



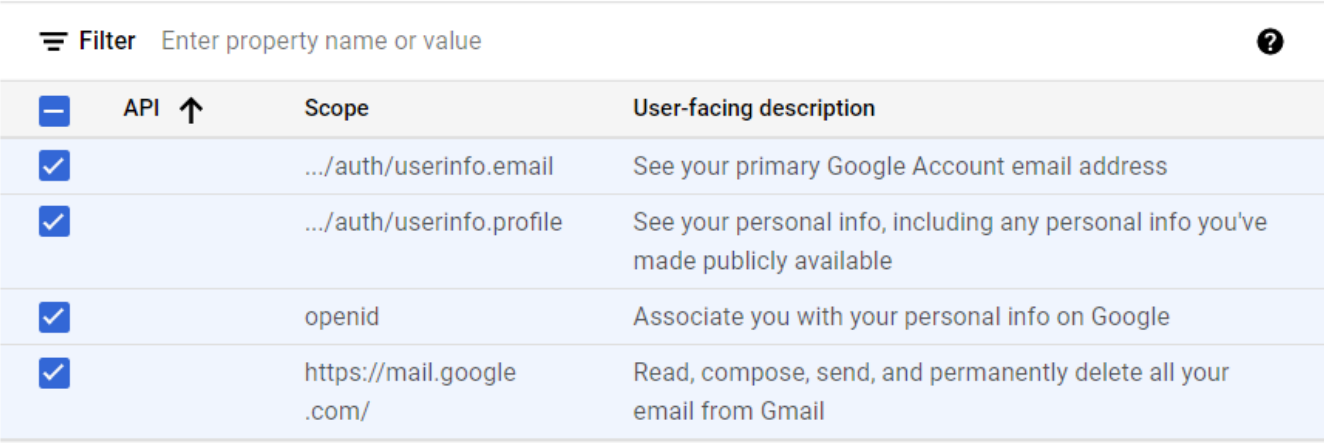
5. Specify the developer's email in the **Developer contact information** section at the bottom of the page and click **Save and continue**.



6. Click **Add or remove scopes**. This will open the **Update selected scopes** dialog on the right.
7. Copy and paste the following text into the **Manually add scopes** field in the bottom part of the dialog and click **Add to table**:

```
openid https://www.googleapis.com/auth/userinfo.email
https://www.googleapis.com/auth/userinfo.profile https://mail.google.com/
```

Note: You can also select scopes manually. The following scopes need to be selected:
- openid
- https://mail.google.com/
- ../auth/userinfo.email
- ../auth/userinfo.profile

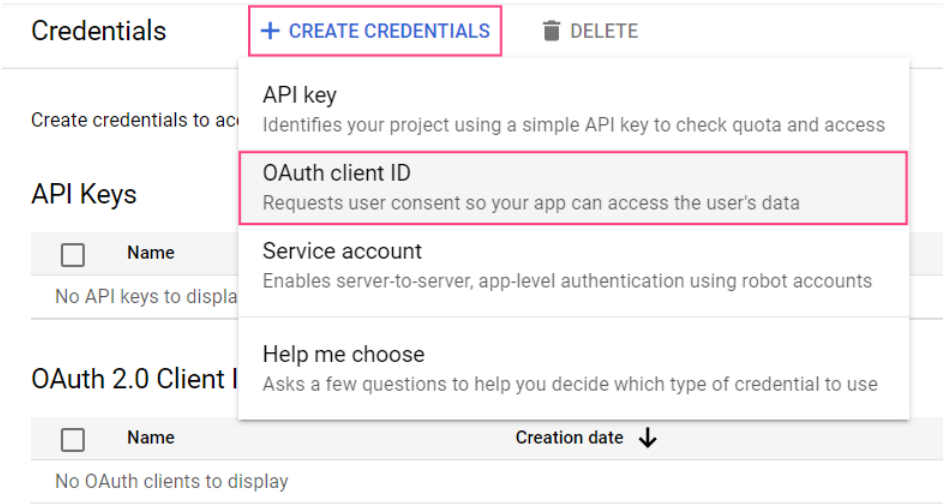


8. Click **Update**. This will close the **Update selected scopes** dialog and display the selected scopes.
9. Click **Save and continue** at the bottom of the screen.
10. Click **Save and continue** to skip the **Test users** page settings and navigate to the **Summary** page.

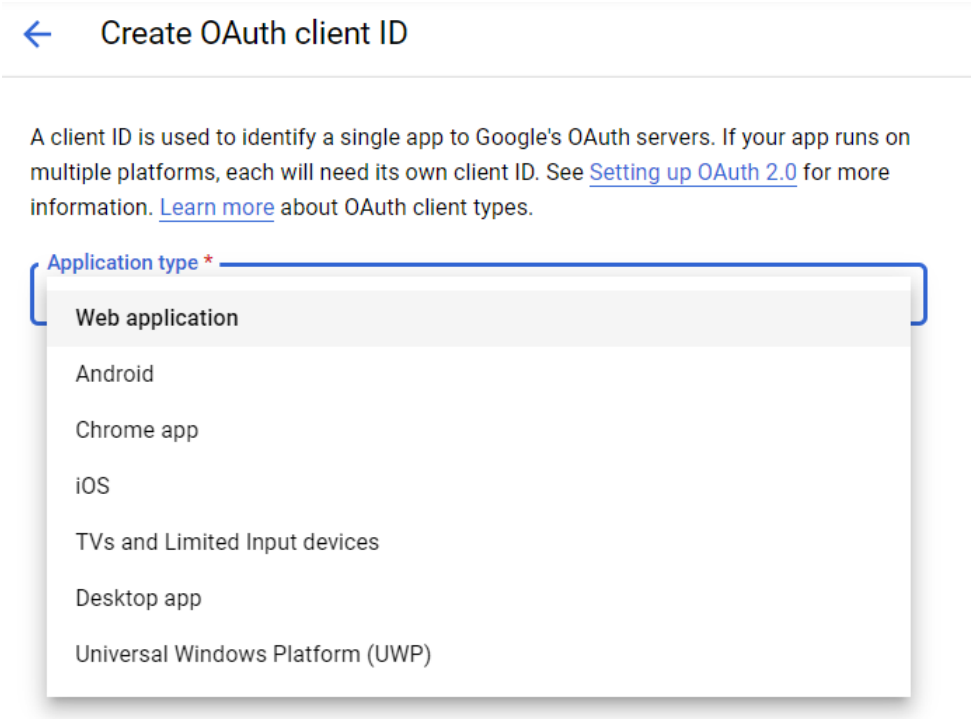
On the **Summary** page, the following is displayed: information about the application, email addresses, and permissions that have been set up.

Creating account credentials

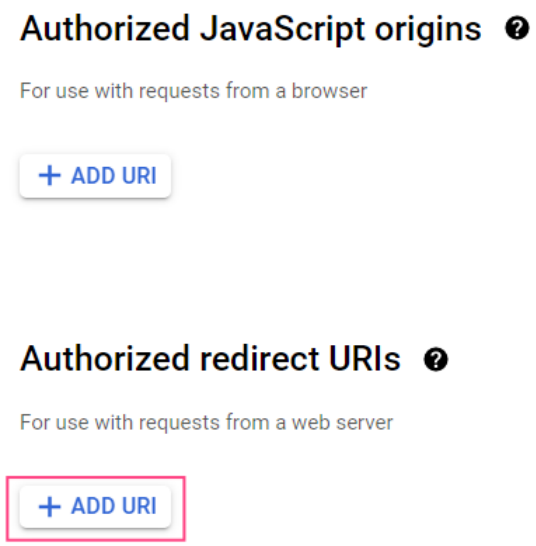
1. Select **Credentials** in the menu on the left side of the screen.
2. Click **+ Create credentials** and select **OAuth client ID**.



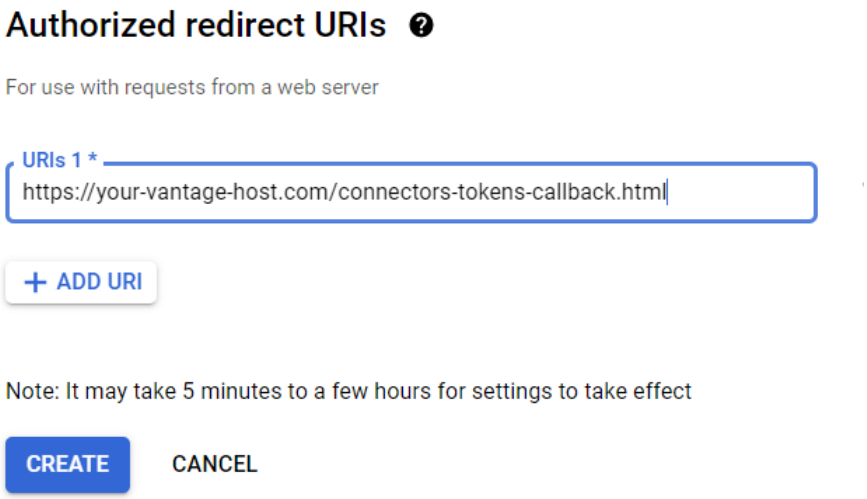
3. Select the Web application type.



4. In the **Authorized redirect URIs** section, select **+ Add URI**.

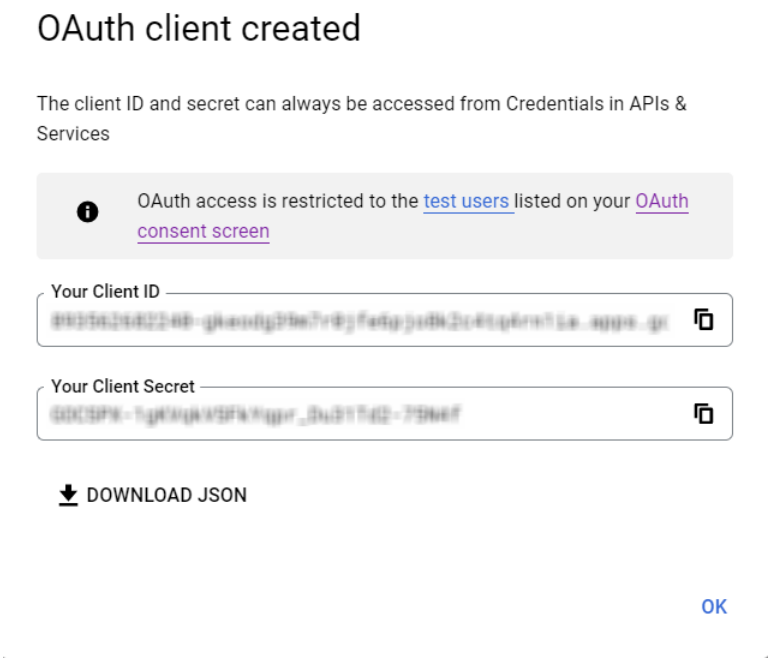


5. In the field that will appear, specify the redirect URI: https://<Vantage host name>/connectors-tokens-callback.html.



6. Click **Create**.

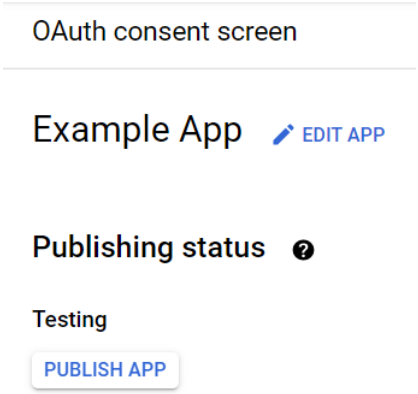
The pop-up dialog box that will appear will contain the Client ID and Client secret values.



This data is required for setting up the tokenmanagement service in Vantage. You can save it immediately or copy it later by navigating to the **APIs & Services** > **Credentials** page in the menu on the left of the screen and selecting the OAuth 2.0 client identifier you have created.

Publishing and verification

The publishing status of the application is displayed in the **APIs & Services** > **OAuth consent screen** section.



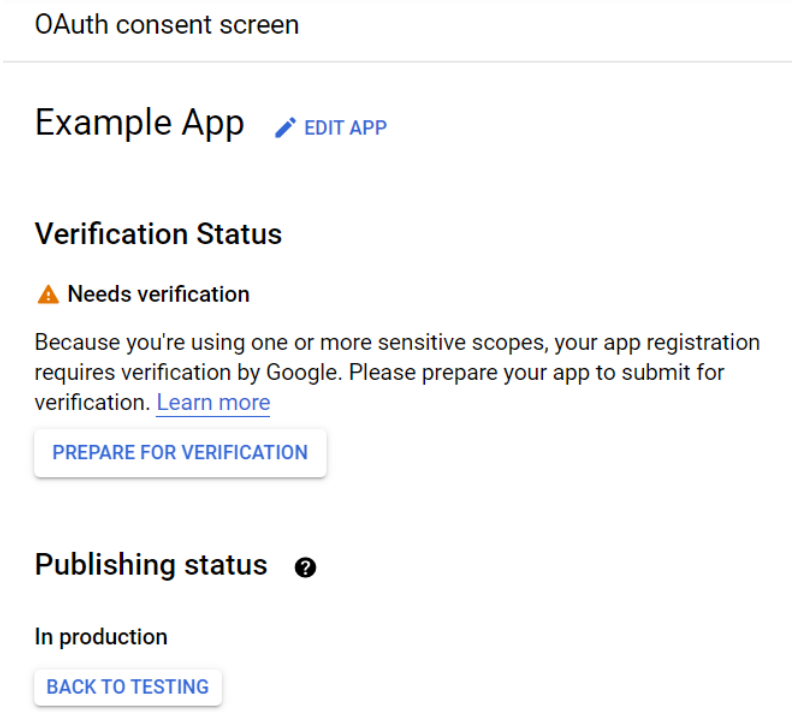
Applications with the **Testing** status are only available to users that have been added to the testers list. Only publishing an application makes it available to any user with a Google account.

Click **Publish app**. The https://mail.google.com/ scope allows the application to access confidential user data, which is why a message saying that the application needs to be verified will be displayed. To verify the application, you will need to provide the following:

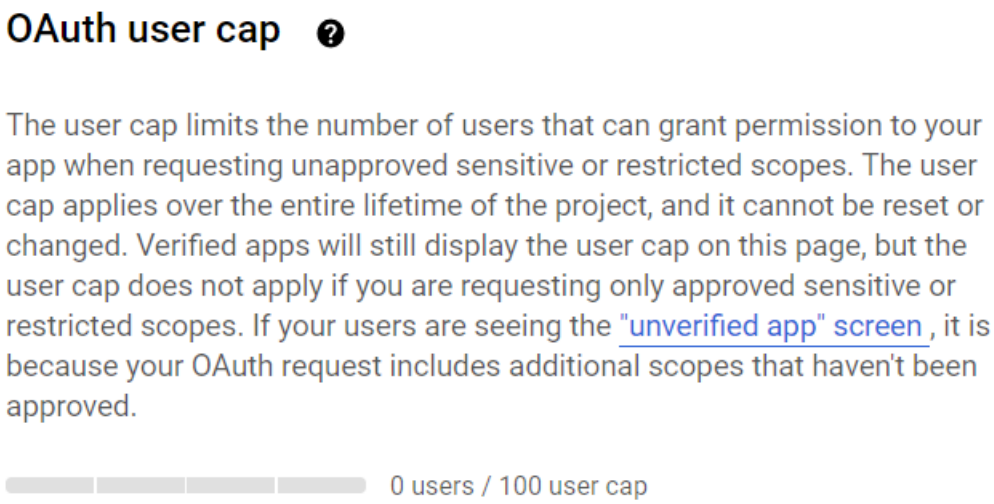
- An official link to the application's Privacy policy;
- A YouTube video demonstrating the stated purpose of obtaining Google user data using the application;
- A text addressed to Google that contains a description of why you require access to confidential user data;
- A full list of all your domains verified in the Google Search Console.

Click **Confirm**. The status of your application will change to **In Production**.

The **Prepare for verification** button will also appear. This button lets you provide all required verification data.



Note: Before your application has been verified, only 100 users are able to use it. The user counter is located in the bottom part of the **OAuth consent screen** section and cannot be reset throughout the projects's lifetime.



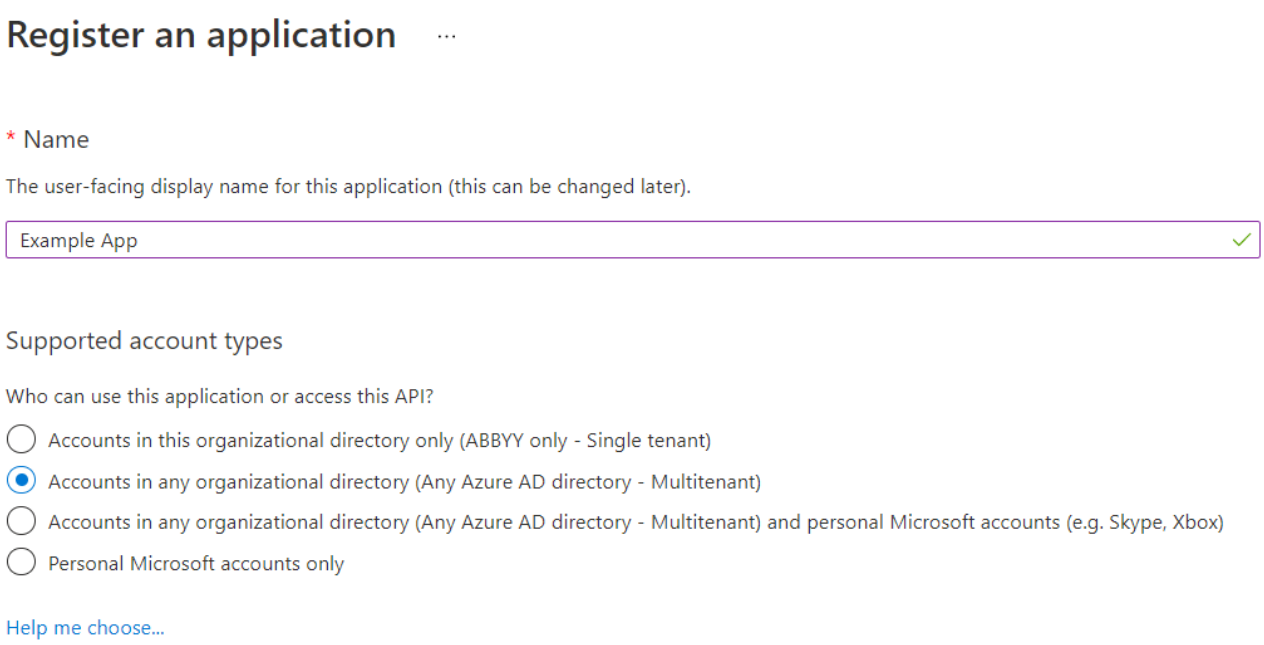
Registering the Application in Microsoft Azure

To create an application, an Azure Active Directory tenant with application registration and editing permissions is required.

You can switch to the correct directory on the Portal settings | Directories + subscriptions page.

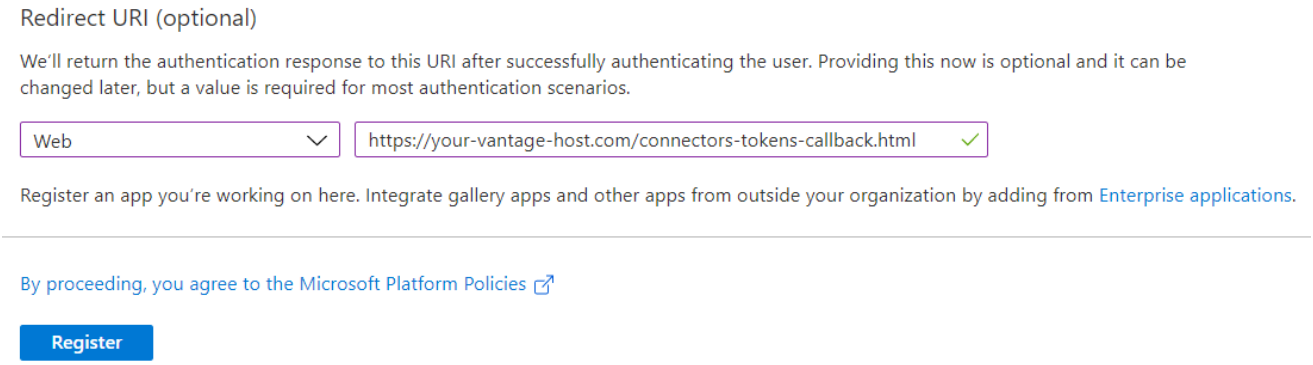
Registering the application

1. Navigate to the App registrations page.
2. Click **New registration**.
3. Specify a name for your application and select the supported account types.



Note: If the Multitenant type is selected for the application, it will be available for users in any Azure AD tenant. Such applications need to be verified, which is only available for Microsoft Partner Network participants. If you are not a participant, select Single tenant, which will only make your app available to users in your own Azure AD tenant.

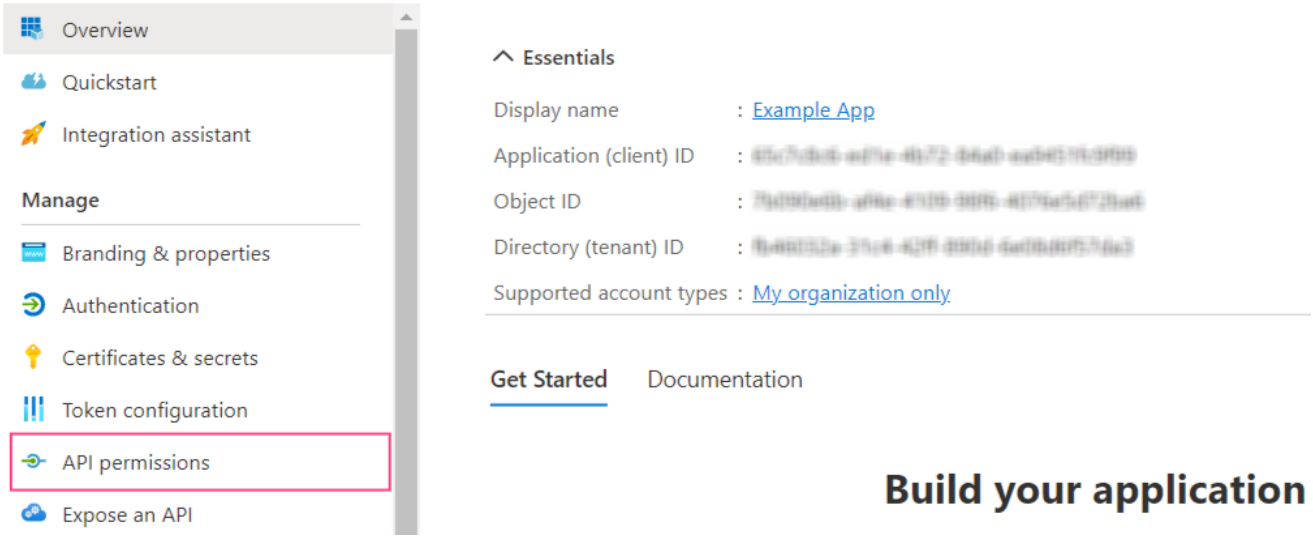
4. In the **Redirect URI** section, select the Web platform and specify the redirect URI: https://<Vantage host name>/connectors-tokens-callback.html.



5. Click **Register**.

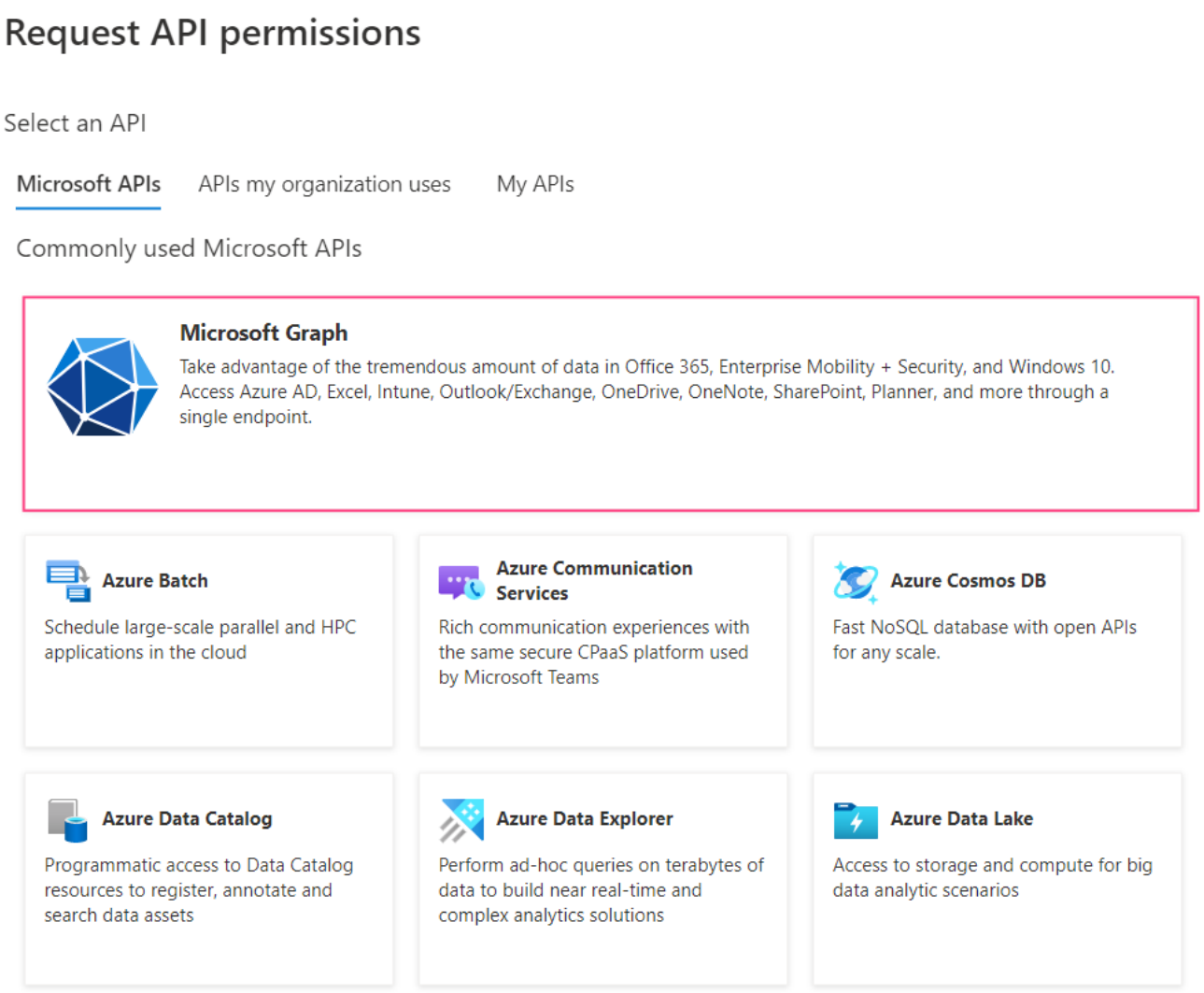
Setting up application permissions

1. Navigate to the **API permissions** tab.

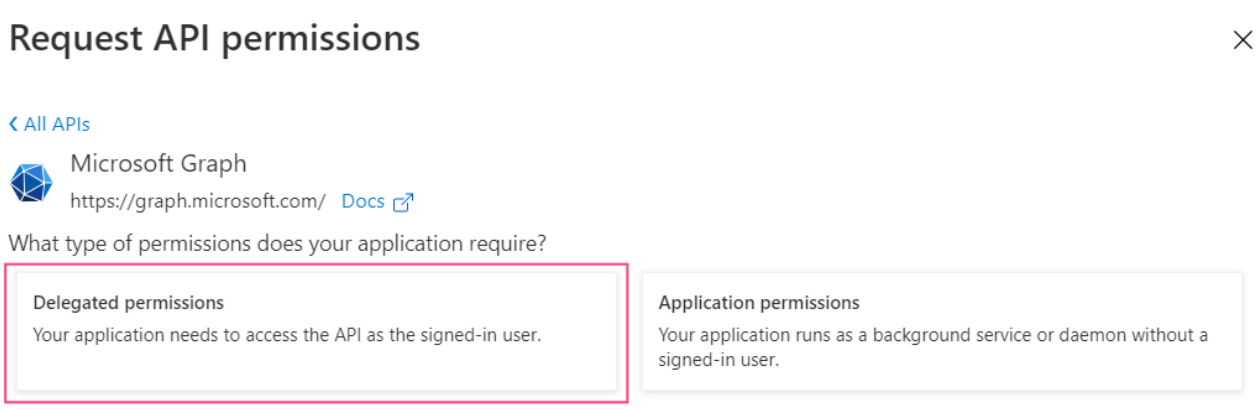


2. Click **Add permission**.

3. In the dialog that will open, select the **Microsoft Graph** section.



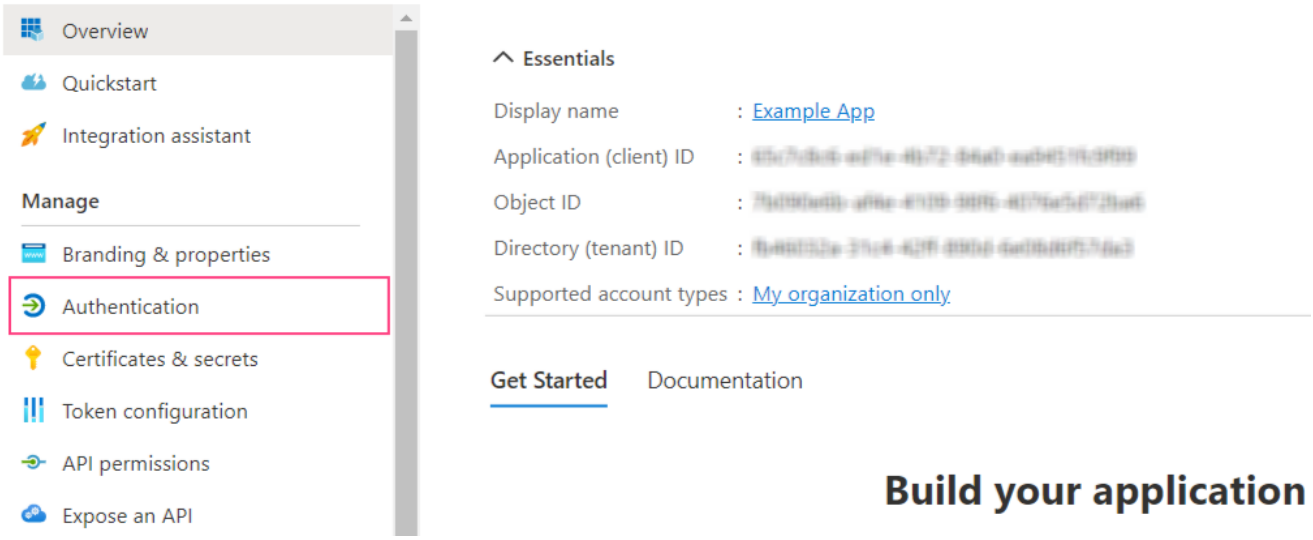
4. Select **Delegated permissions**.



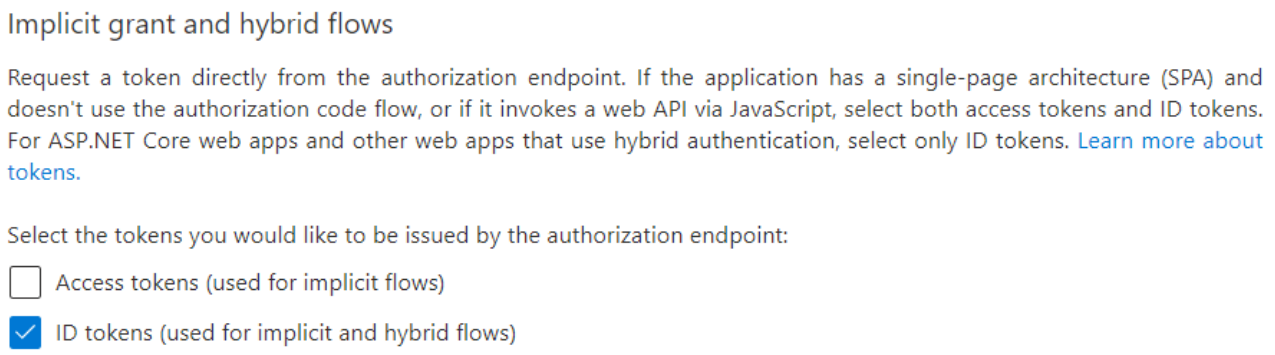
5. Add the following permissions:
    - email
    - IMAP.AccessAsUser.All
    - offline_access
    - openid
    - profile
6. Click **Add permissions**. This will close the dialog and display the selected permissions.

Creating client secrets

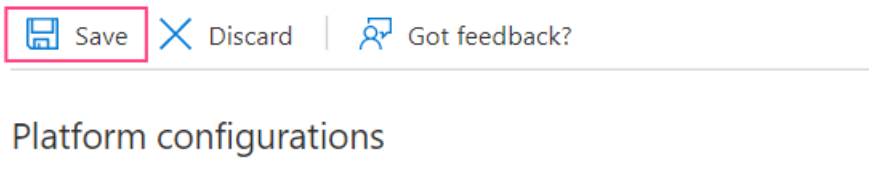
1. Navigate to the **Authentication** tab.



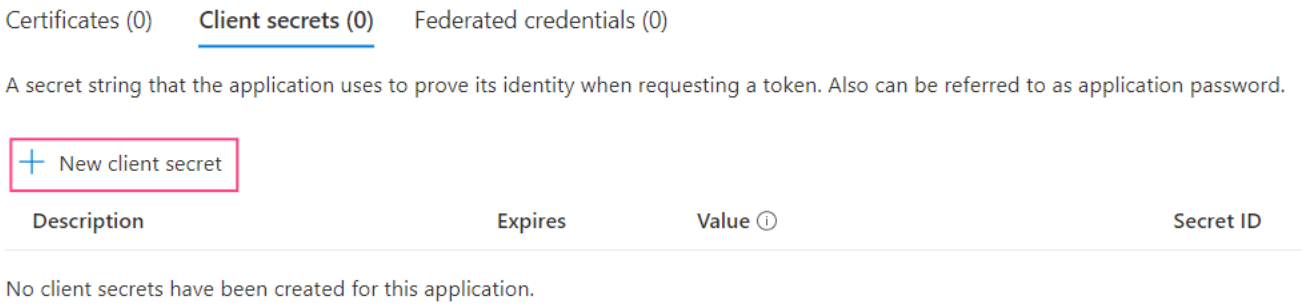
2. In the **Implicit grant and hybrid flows** section, mark **ID tokens (used for implicit and hybrid flows)**.



3. Click **Save** at the top of the screen.



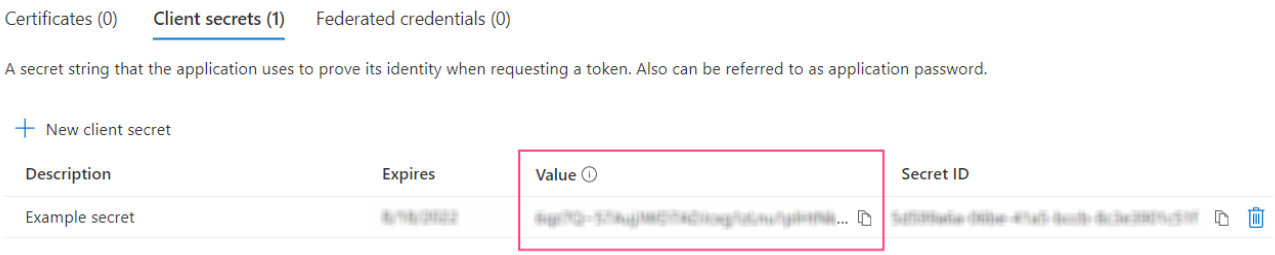
4. Navigate to the **Certificates & secrets** tab and click **New client secret**.



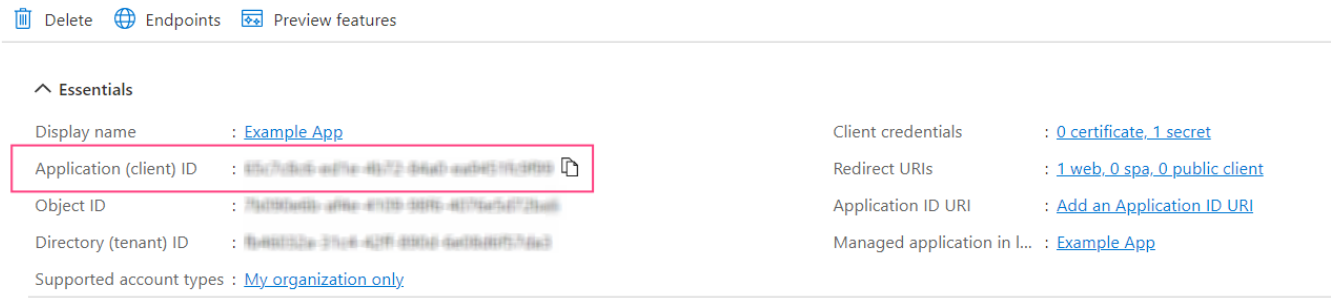
5. In the dialog box that will open, specify a name for the client secret and an expiration date.

Note: The maximum expiration date is 24 months.

6. Click **Add**. This will close the dialog and display information about your new client secret. It is important that you copy and save the **Value**, since you will not be able to access it again once you close the page. **Value** is required when configuring the tokenmanagement service in Vantage.



You will also need a client identifier, which can be copied from the **Application (client) ID** field in the **Overview** tab. The copy icon will appear once you hover the mouse cursor over the value of the identifier.

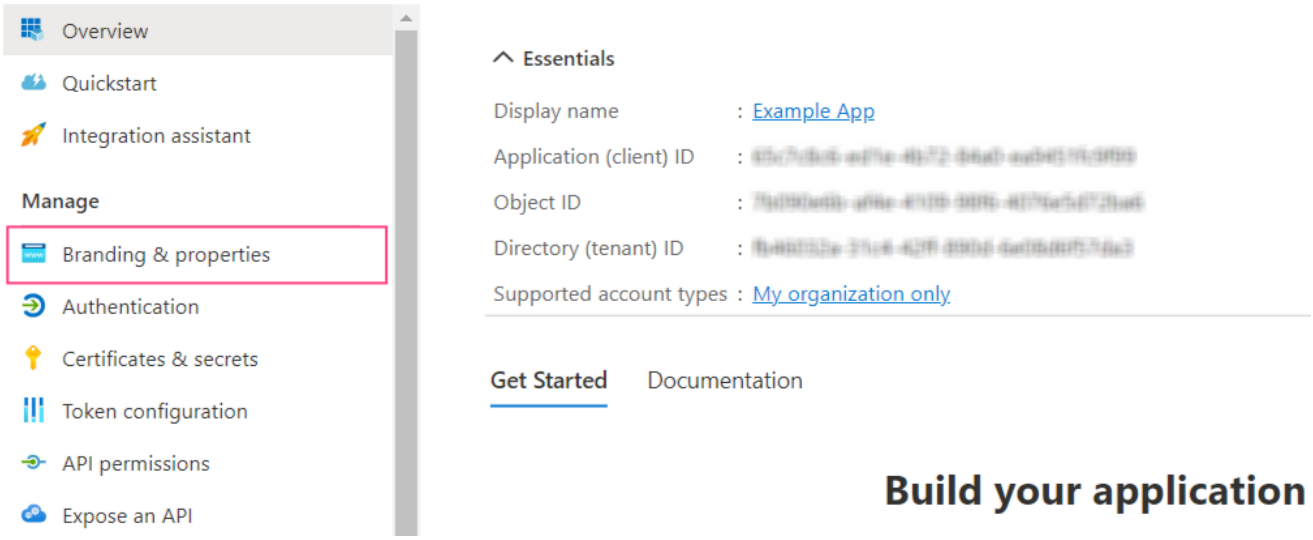


Verifying the application

To make the application available to users from any Azure AD tenant, verification is required. Verification is not required if accounts from a single Azure AD tenant are used.

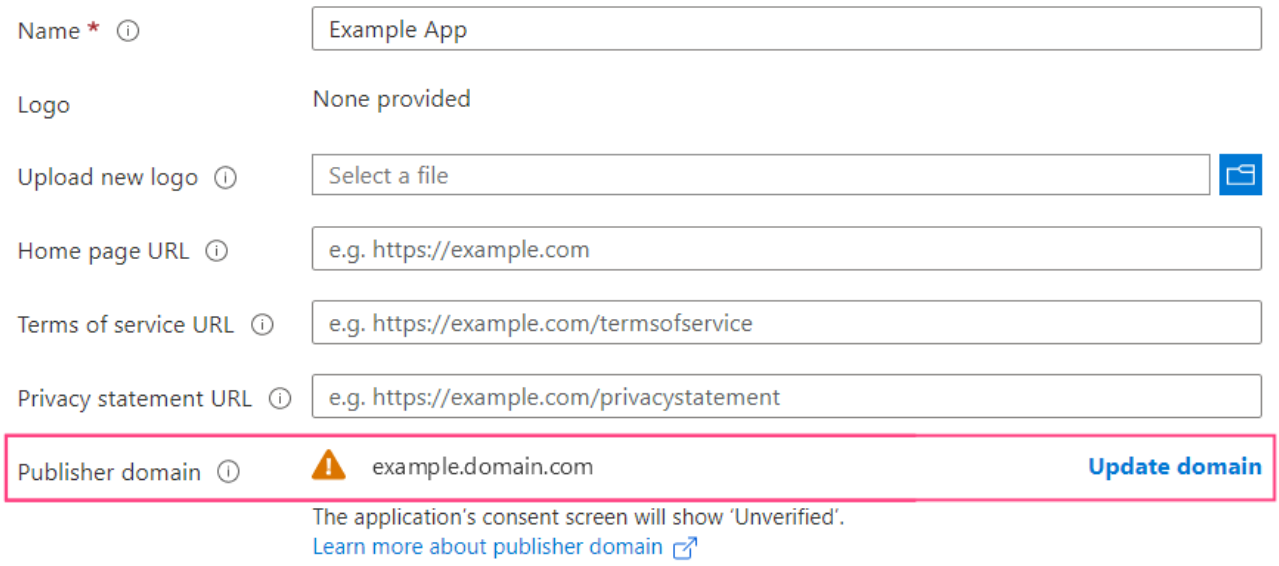
Only Microsoft Partner Network participants can undergo verification.

1. Navigate to the **Branding & properties** tab.



2. Verify that the domain is specified in the **Publisher domain** field. If required, configure your domain by clicking **Configure a domain**.

   The warning icon displayed next to the domain name means that an application with the specified domain cannot be verified. Click **Update domain** to specify a different valid domain related to the Azure Active Directory tenant. Alternatively, verify a new domain.



3. In the **Publisher verification** section, specify your MPN ID and click **Verify and save**.

Note: If you do not have the required permissions to add an MPN ID, verify that all publisher verification requirements are satisfied.

Once your verification is successful, the appropriate icon will be displayed next to the **Publisher display name** field.

Passing Credentials to Consul

If Vantage is already installed, you need to use Consul to manually enter account credentials generated for Microsoft and/or Google email service authentication.

Features specific to the OAuth 2.0 protocol are listed in the TokenManagement service.

Note: Before setting up, verify that the kubectl command line tool is installed and that you are connected to the Kubernetes cluster.

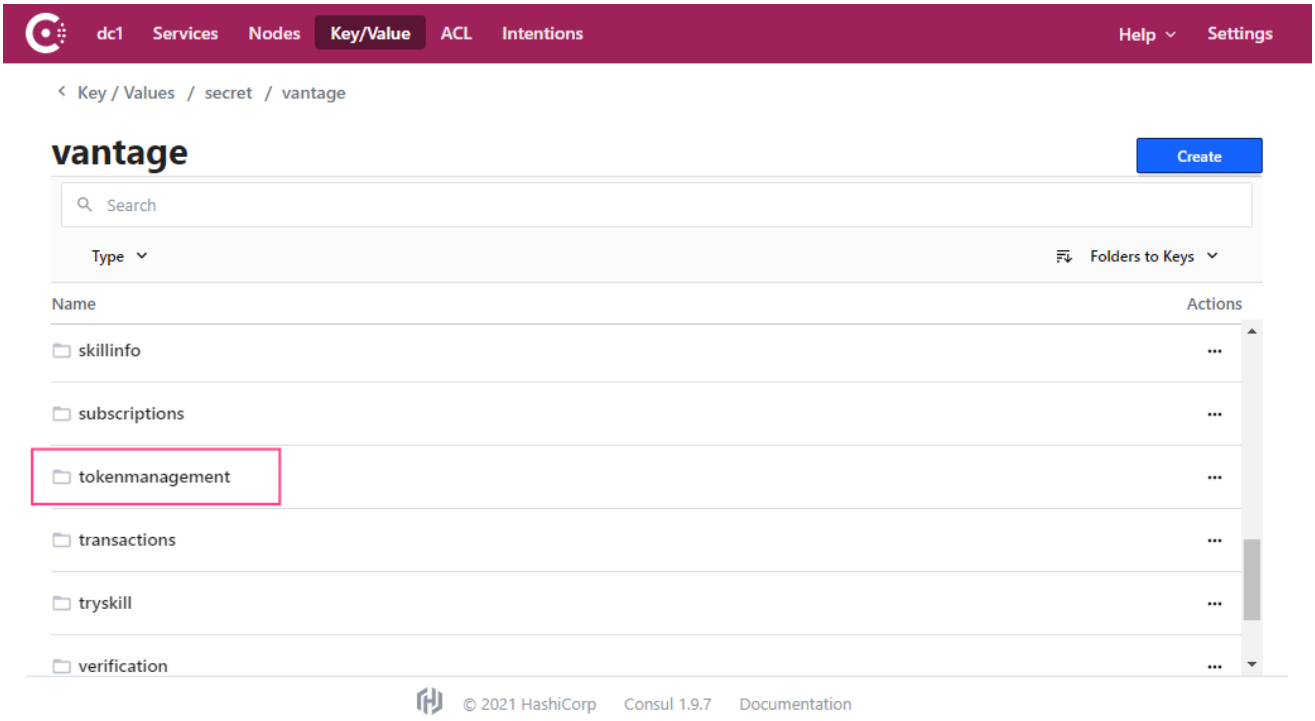
1. Get access to the Consul web interface by running the following command:

   ```
   kubectl port-forward -n infrastructure service/consul-ui 8500:80
   ```

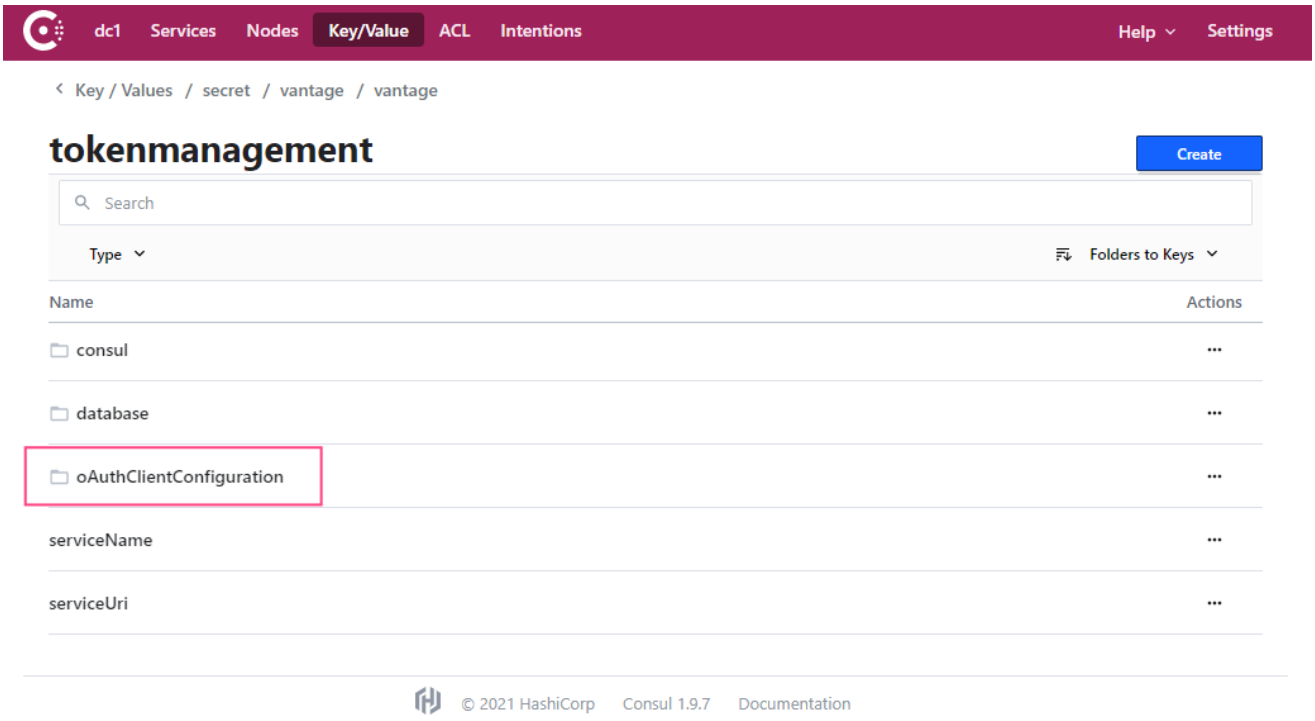
   Next, navigate to http://localhost:8500/ui/dc1/kv/secret/.

2. In the **Key/Value** tab that will open, select the appropriate Vantage deployment scope. Then, select the **vantage** project.

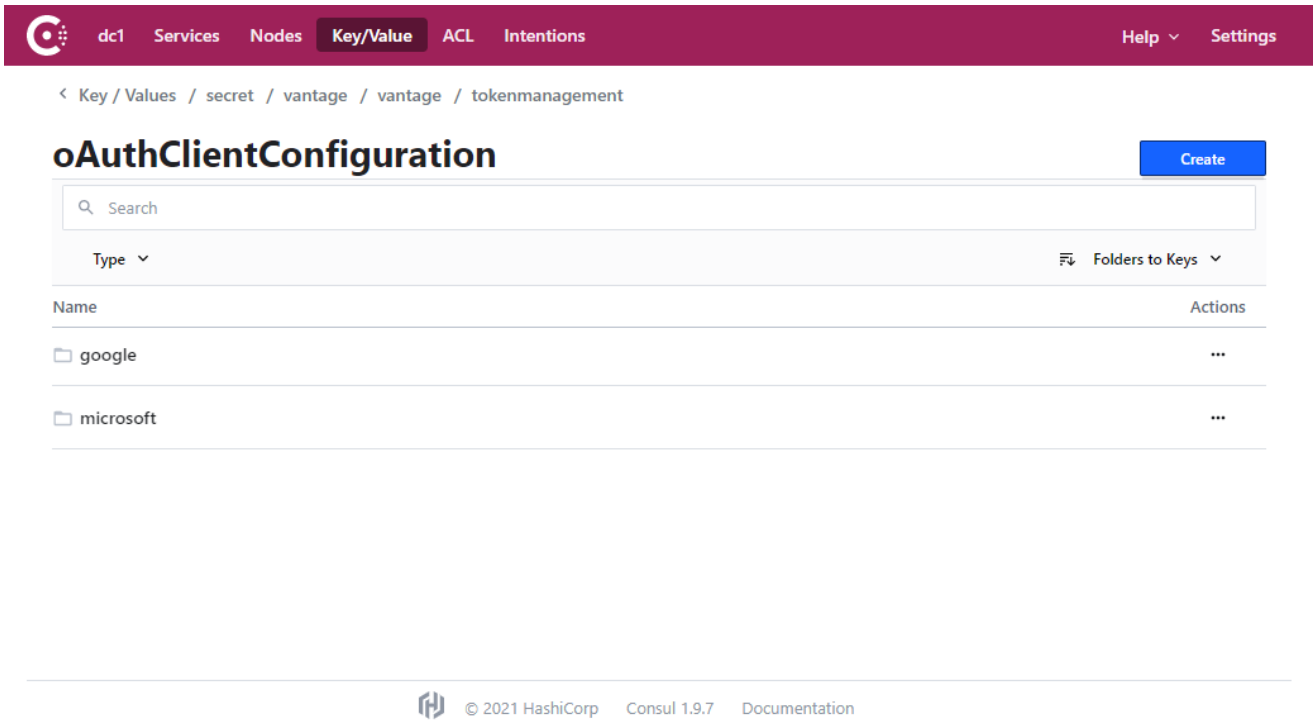
3. Select the **tokenmanagement** service.



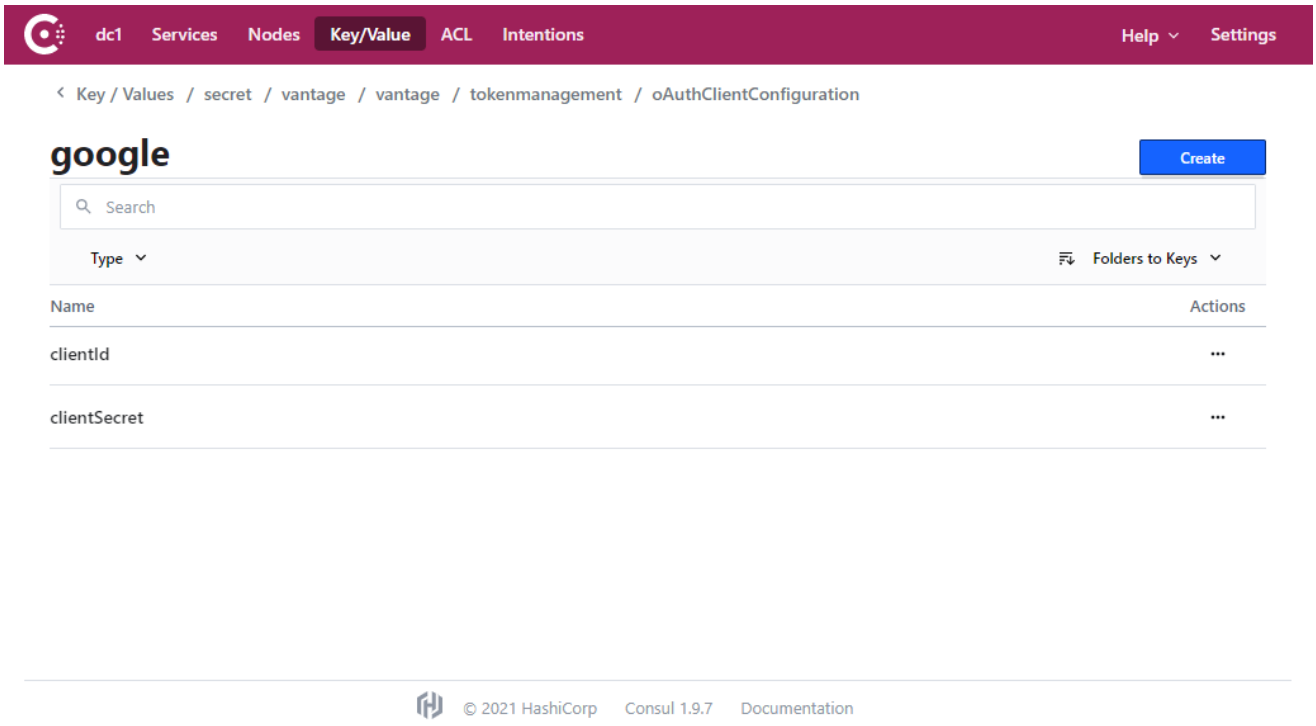
4. Navigate to the **oAuthClientConfiguration** section.



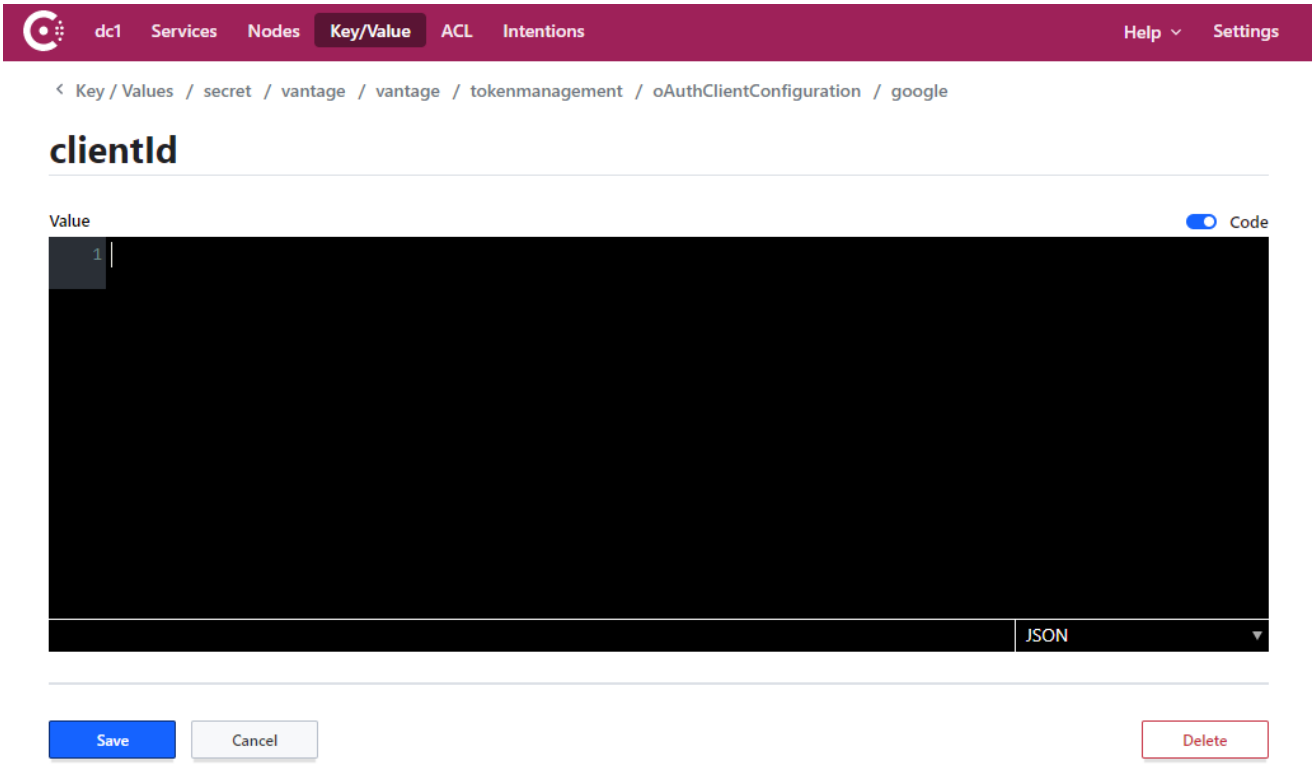
5. Select the service for which you want to specify user data.



6. Select the **clientId** key.



7. Copy and paste the Client ID value you saved earlier to the entry field and click **Save**.



8. Repeat steps 6 and 7 for the **clientSecret** key.

If required, repeat steps 5 through 8 for a different email service.

9. Restart the **tokenmanagement** service by running the following command:

```
namespace={your-namespace}
label=tokenmanagement
kubectl -n $namespace rollout restart $(kubectl -n $namespace get deployments -l
app.kubernetes.io/component=$label -o name)
```

Updating Client secret

The Client secret value is used for serverside client identification and constitutes confidential information. For security purposes, data like this should periodically be updated. Some services like Azure Active Directory limit the validity period for such data.

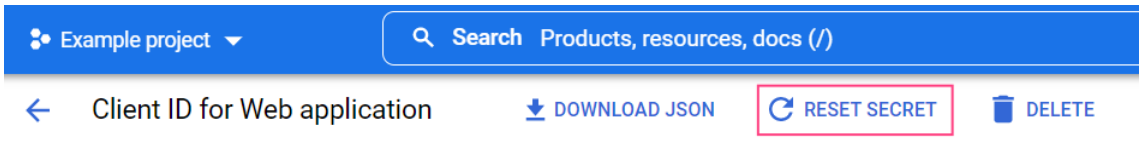
Once a new Client secret has been created, the value of the corresponding Consul key should also be updated.

Note: Once Client secret has been updated, users will need to set up connections to their email service in the Input activity of the Document Skill from scratch. Otherwise, Vantage will not be able to connect to the mailbox and import emails from it.

Updating Client secret in Google

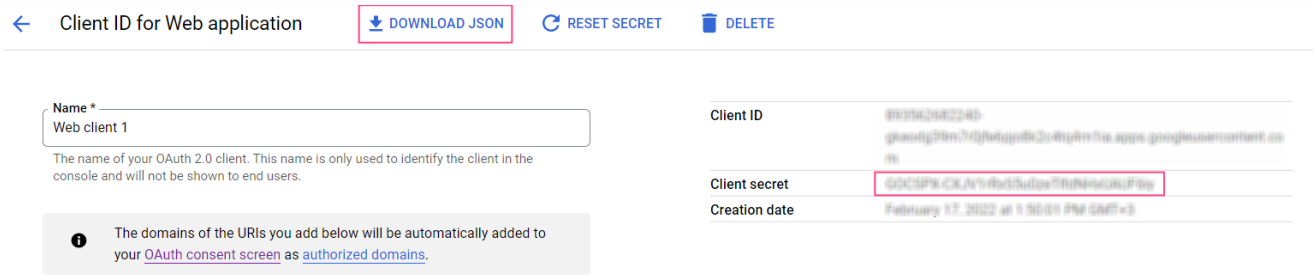
1. Navigate to the Google Cloud Console and select the appropriate project.
2. In the menu on the left, select **APIs & Services** > **Credentials**.
3. In the **OAuth 2.0 Client IDs** section, select the identifier used to authenticate when connecting to the IMAP server.

4. Click **Reset secret**.



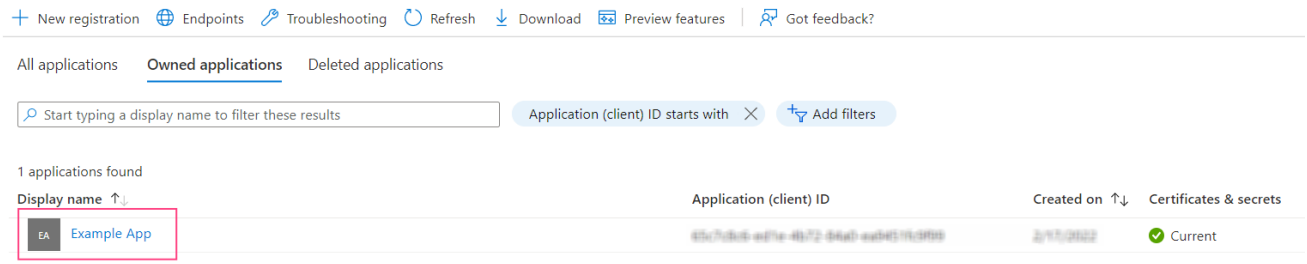
5. Click **Reset** in the pop-up dialog box. This will update the Client secret value and recall its previous value.

6. Download the JSON file containing the credentials. Alternatively, copy the Client secret value from the right side of the screen.



Updating Client secret in Microsoft Azure

1. Navigate to the App registrations page and select the application used for authentication using the IMAP server.



2. Navigate to the **Certificates & secrets** tab and click **New client secret**.

3. In the dialog box that will open, specify a name for the client secret and its expiration date.

4. Click **Add**. This will close the dialog and display information about the new client secret. It is important that you copy and save the **Value**, since you will not be able to access it again once you close the page.

5. If the current client secret has not expired yet, you can delete it in order to only be able to use the new client secret to identify the client.

Updating Client secret in Consul

Follow the steps listed in the Passing Credentials to Consul section, omitting steps 6 and 7 (copying the **clientId** value).

Monitoring and Administration

As a system administrator, you are tasked with monitoring ABBYY Vantage at all times, managing it, discovering any errors that may occur during document processing, as well as the causes of such errors.

You can do this by using:

- Vantage logs.
- The built-in **Skill Monitor** service, which collects statistics for existing Vantage skills and provides detailed information regarding completed and ongoing transactions. This service also lets you get transaction event information required by technical support.
- A range of more advanced third-party services which allows you to: monitor internal Vantage processes, monitor specific workflows, analyze collected data to further fine-tune and optimize document processing, collect and analyze logs.

Logs

To get ABBYY Vantage logs, do the following:

1. Navigate to SMB share **\\X.X.X.X\cluster_volumes**:

```
smb:
  host: X.X.X.X
  share_name: cluster_volumes
```

2. Search it for the word "logs".

Grafana

Grafana is a tool for visualizing, monitoring, and analyzing data. Grafana is not included with the installation of ABBYY Vantage and has to be set up separately by doing one of the following:

- Installing Grafana Helm Chart.
- Installing Grafana Operator.

In either case, ensure that Helm (Kubernetes package manager) and OpenShift Origin Client tools (OC tools) are both installed.

Installing Grafana using Helm Chart

To install Grafana using Helm Chart, do the following:

1. Upload the **values.openshift-persistence-off.yaml** file found in the installation package.
2. Modify the following file parameters: domain name (Fully Qualified Domain Name), Grafana administrator login and password.
3. Install Grafana Chart by running the following command:

```
helm upgrade -i -n monitoring grafana grafana/grafana -f values.openshift-persistence-off.yaml
```

4. Get a token for your account by running the following command:

```
oc sa get-token grafana -n monitoring
```

5. Look for configMap in the namespace and update **Bearer some_token** in **datasource.yaml** by adding the value of the received token (**Bearer *your*_token**).

6. Restart Grafana.

Once you have carried out the above steps, preset dashboards for various data sources will be displayed in the integrated Grafana interface in ABBYY Vantage. Setting up filters and variables will let you view information about the system using various dashboard graphs.

Installing Grafana Operator

You can install Grafana Operator in the **Monitoring** namespace using instructions from the official RedHat website. For samples of kubernetes manifests for CDR (Custom Resource Definition) tables, see the **grafana-operator-manifests.yaml** file in the installation package.

To use ABBYY dashboards, launch the **migrate_dashboards.sh** scenario and verify that the kubernetes manifests generated in file **grafanaDashboards.yaml** are correct.

Grafana Overview Dashboard

Preset dashboards for various data sources are displayed in the integrated Grafana interface in ABBYY Vantage. The main dashboard called Overview displays a general overview of the status of ABBYY Vantage. This dashboard contains the following graphs:

- Base measurements such as **Failed Nodes**, **HTTP Error Rate**, as well as panels for monitoring the status of the Redis cluster.
- Graphs for services such as **Platform Services Request Rate**, **Vantage Services Response Time**, etc.
- Graphs for workers such as **OCR Workers**, **Training Workers**, **OCR Workers Processing Time**, etc.

You can switch between different types of data and graphs displayed on the dashboard by setting up the filters and variables accordingly.

EULA and Privacy Policy Links

You can find the End User License Agreement document in the ABBYY Vantage distributive by the following path: `/legal/eula.txt`.

Data Use and Protection Policy is available at official ABBYY site.