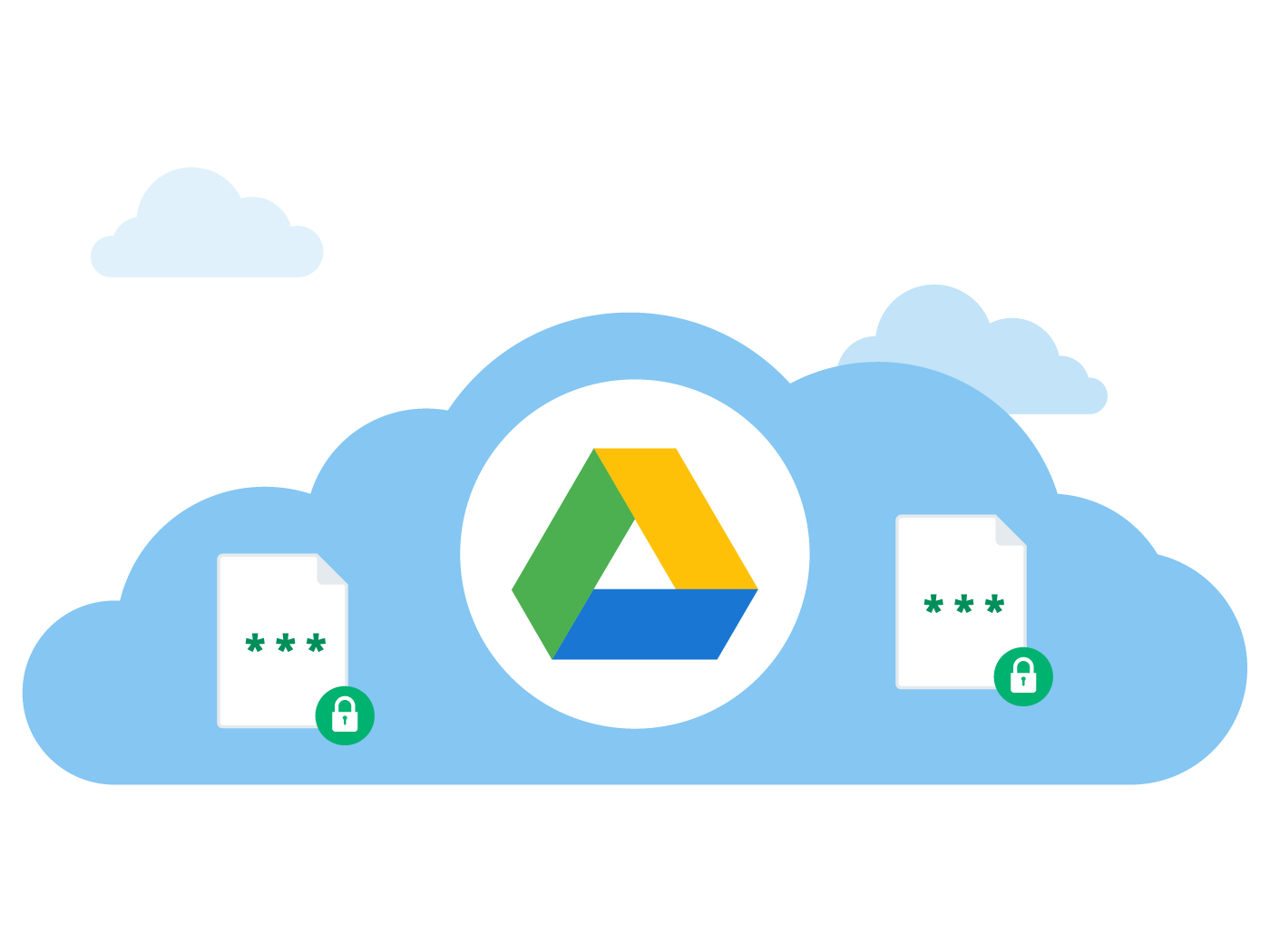
Google Drive with Pentaho Data Integration

Introduction

Google Drive is cloud-based storage, which means no matter where you are or what computer you use to save a file to Google Drive, it will show up anywhere you can access Google Drive. So let’s say you are working on a lesson plan at home and you upload the file to Google Drive, then you get to work the next day and open Google Drive; the file will be there.



Pre-requisites

The following pre-requisites are required:

* Pentaho Data Integration and above with optional BA Pentaho Server.
* Google Drive Account
* oAuth Credentials
* Transformations / Jobs / Dataset

Step 1: Enable Google Drive API

The Drive platform gives you a group of APIs along with client libraries, language-specific examples, and documentation to help you develop apps that integrate with Drive.

The core functionality of Drive apps is to download and upload files in Google Drive. However, the Drive platform provides a lot more than just storage. This page describes some of that functionality and points you to resources for building it into your app. Enabling the Google Drive API can be achieved via the following steps:

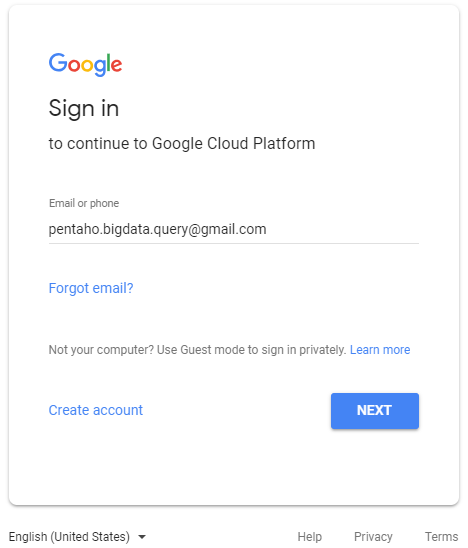
1. Connect with the browser to:

https://console.developers.google.com/apis/library

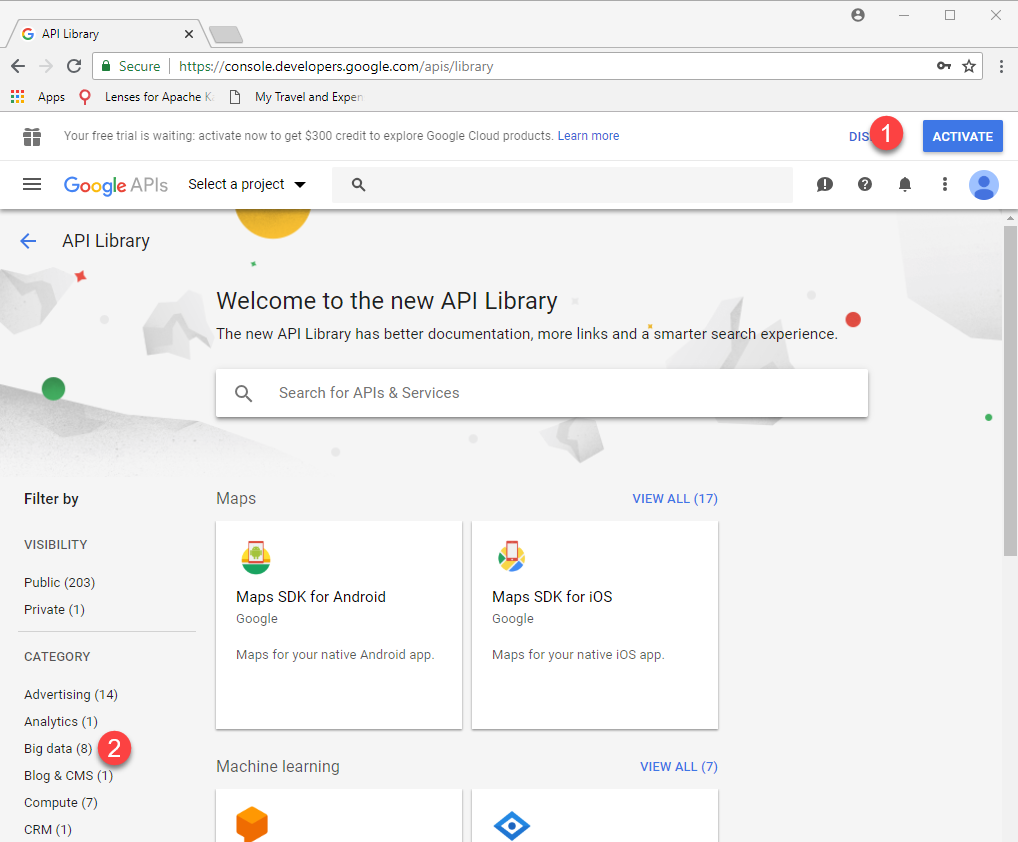
1. Login with the following Google Account that will be used to connect:

**eMail:** pentaho.bigdata.query@gmail.com

**Password:** BigQuery01!



The Google API Library main page will be shown:



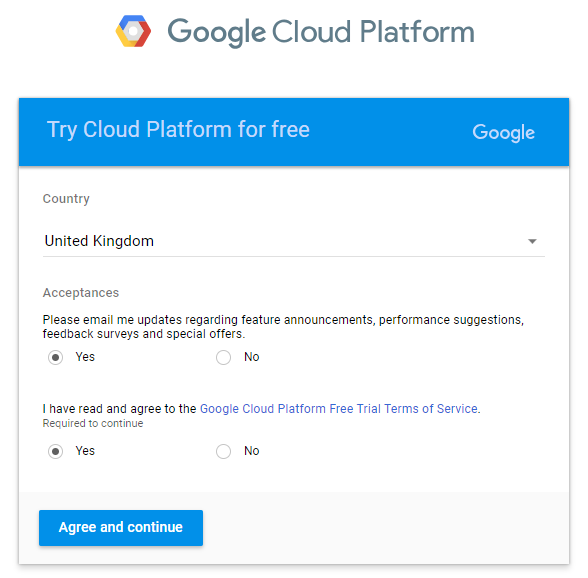
*If this is the first time you’re setting up the account you will need to ACTIVATE.*

*You will require your credit card details. The account will not be charged unless you manually upgrade to a paid account.*

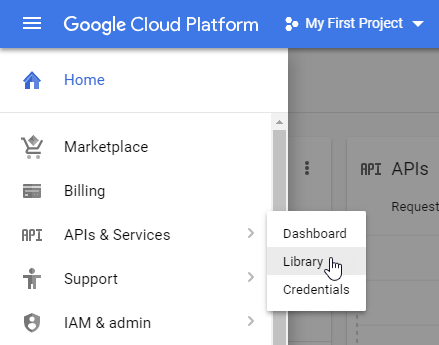
For further details of the Google Free Tier:

https://cloud.google.com/free/

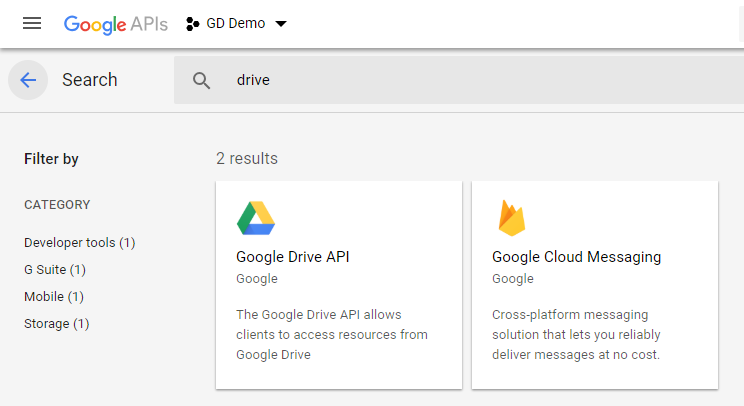
https://cloud.google.com/bigquery/pricing#free



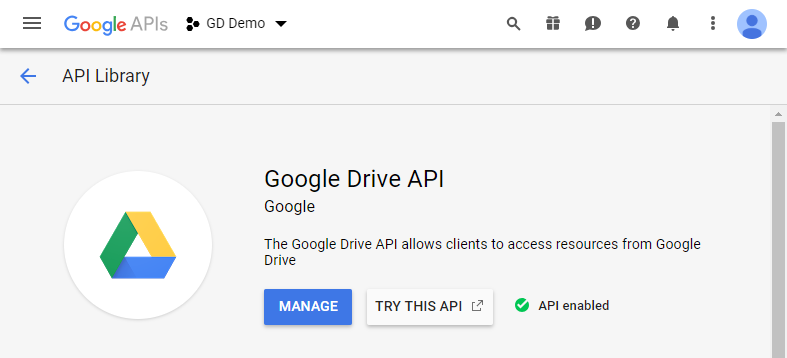
1. Accept the Terms
2. From the Account Dashboard, select APIs & Services > Library.



1. Click on Storage category or search for Drive



1. Click on “Google Drive API”
2. Ensure that the Google drive API is enabled.

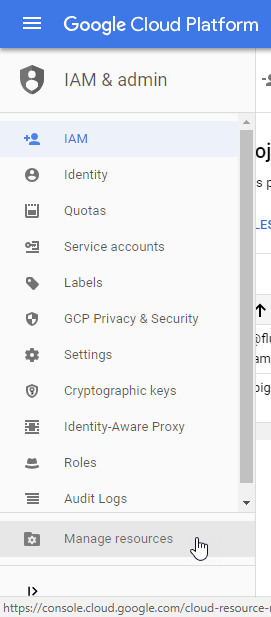


Step 2: Creating and Managing a Google Drive Project

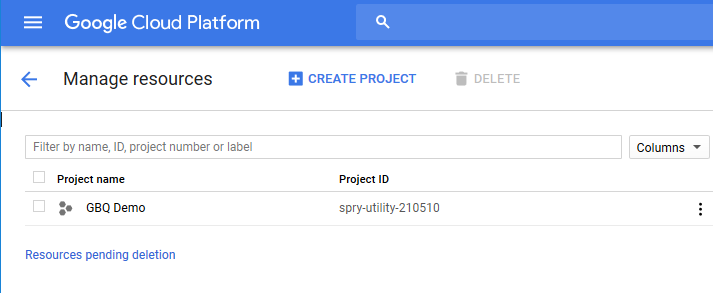
Google Cloud Platform projects form the basis for creating, enabling, and using all GCP services including managing APIs, enabling billing, adding and removing collaborators, and managing permissions for GCP resources.

To create a new project:

1. Go to the GCP Console Manage resources page.



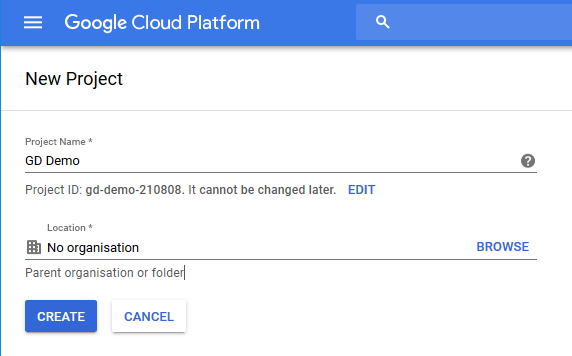
1. Select the option Manage Resources
2. On the drop-down at the top of the page, select the organization in which you want to create a project. (Not applicable as Individual account)
3. Click Create Project



1. In the New Project window that appears, enter:

GD Demo

as the project name and select a billing account as applicable.



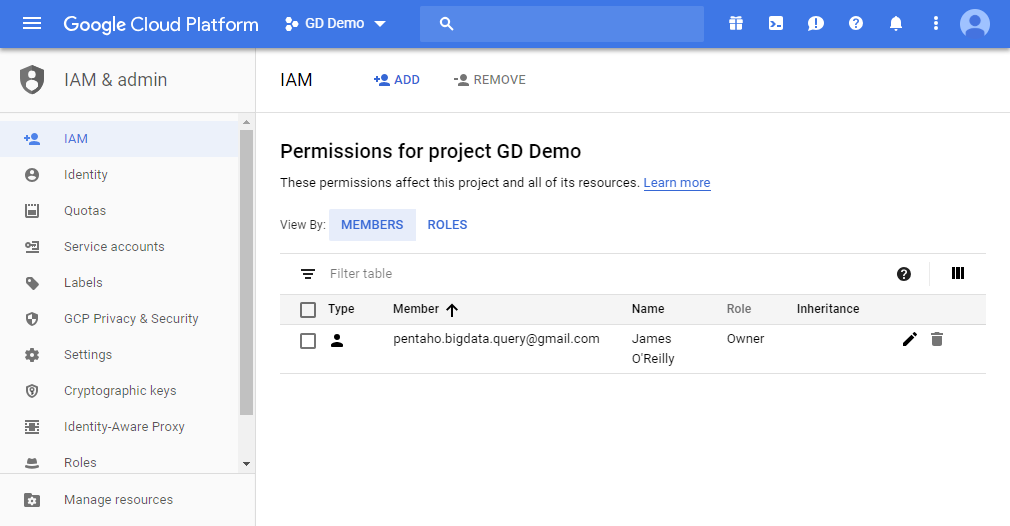
If you want to add the project to a folder, enter the folder name in the Location box.

1. When you're finished entering new project details, click Create.

Identity and Access Management

Google Cloud Platform offers Identity and Access Management (IAM), which lets you give more granular access to specific Google Cloud Platform resources and prevents unwanted access to other resources. IAM lets you adopt the security principle of least privilege, so you grant only the necessary access to your resources.

IAM lets you control who (users) has what access (roles) to which resources by setting IAM policies. IAM policies grant specific role(s) to a user giving the user certain permissions.



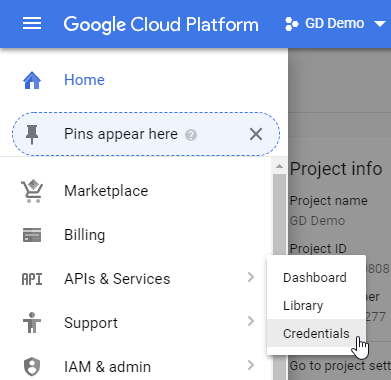
OAuth 2.0 Authentication

Google APIs use the OAuth 2.0 protocol for authentication and authorization. Google supports common OAuth 2.0 scenarios such as those for web server, installed, and client-side applications.

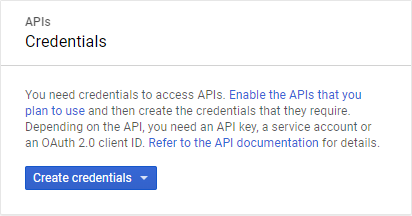
To begin, obtain OAuth 2.0 client credentials from the Google API Console. Then your client application requests an access token from the Google Authorization Server, extracts a token from the response, and sends the token to the Google API that you want to access.

To create an OAuth client ID:

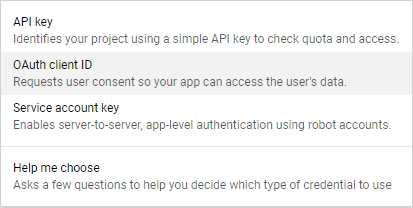
1. In the Google Cloud Platform Console, select APIs & Services > Credentials



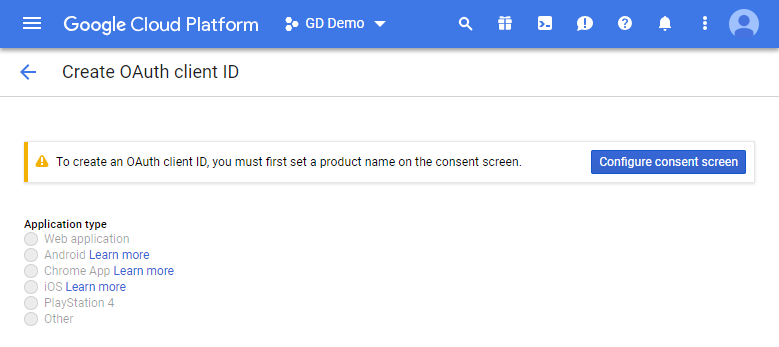
1. Click on Create Credentials



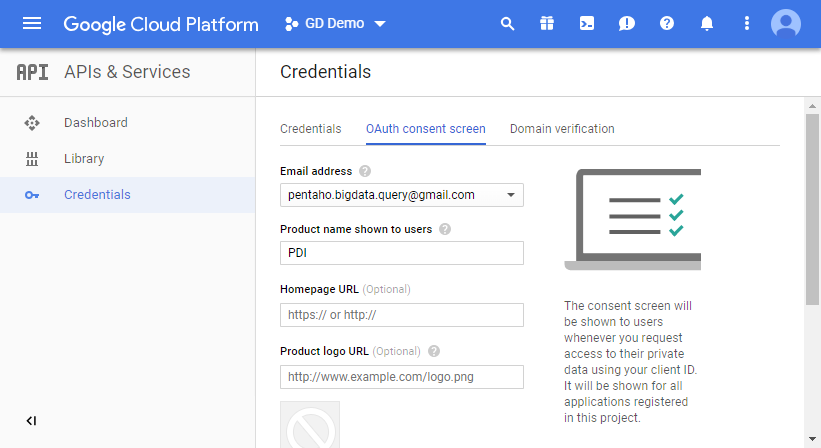
1. Select OAuth client ID



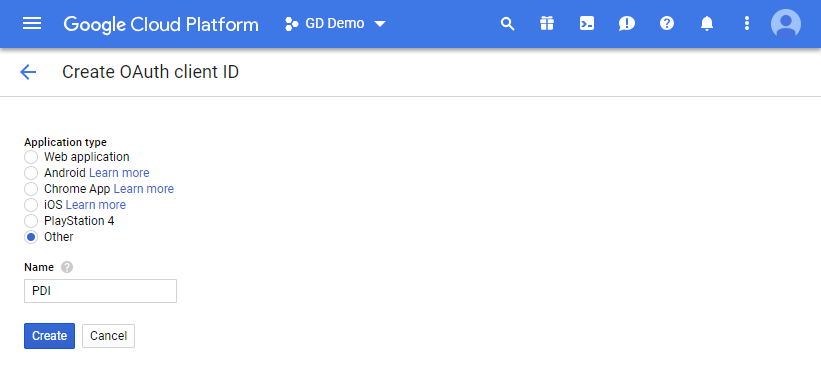
1. Click on Configure consent screen.



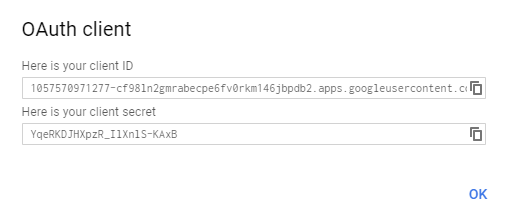
1. Enter a Product Name – PDI



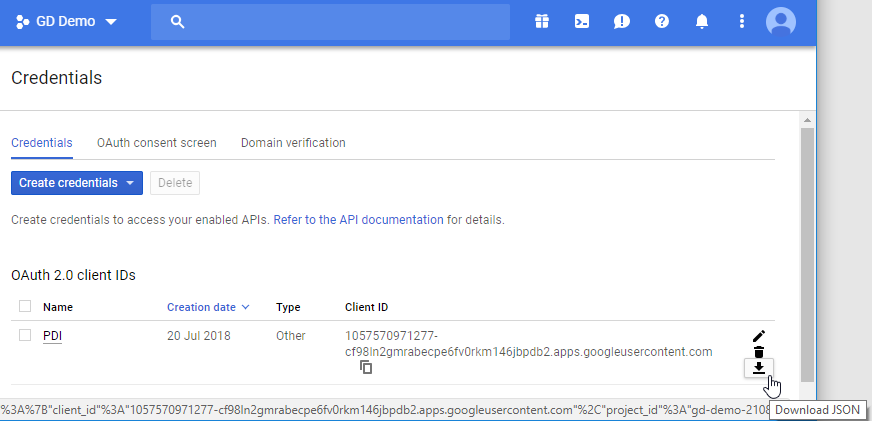
1. Select “Other” as option, enter the Product Name then click on “Create “.



1. The client ID and secret key are created.



1. Click OK
2. Download the client\_secret\_xxx.json key



Step 3: Pentaho Data Integration and Google Drive

Before you can access your Google drive in Pentaho Data Integration:

1. Rename you file to: client\_secret.json
2. Copy your Google client\_secret.json file into (The Google Drive option will not appear as a Location until you copy the client\_secret.json file into the credentials directory and restart)

Spoon:

data-integration/plugins/pentaho-googledrive-vfs/credentials

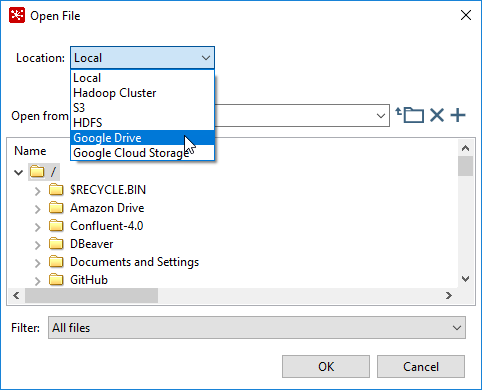
directory, and restart spoon.

Pentaho Server:

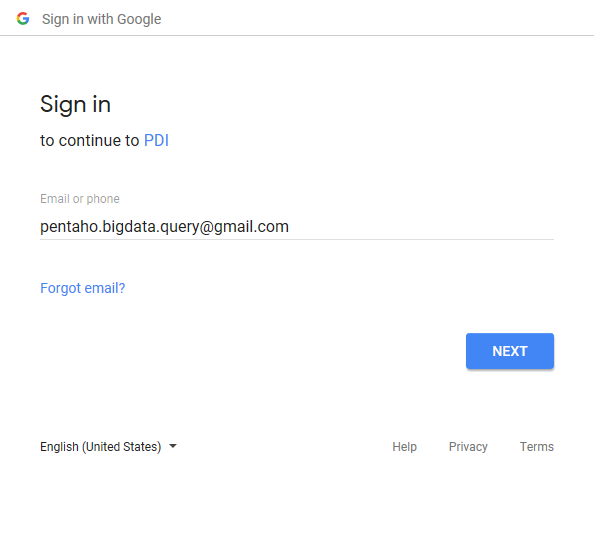
pentaho-server/pentaho-solutions/system/kettle/plugins/pentaho-googledrive-vfs/credentials

directory and restart the server

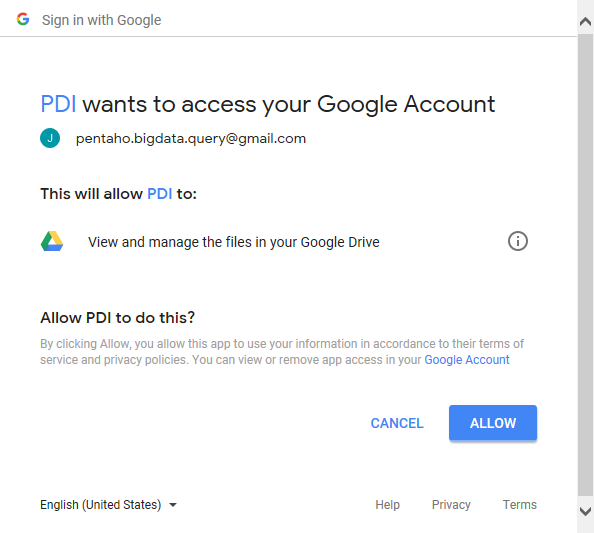
1. In Spoon, Select File > Open URL
2. Select Google Drive from the drop-down box and set the filter for All files



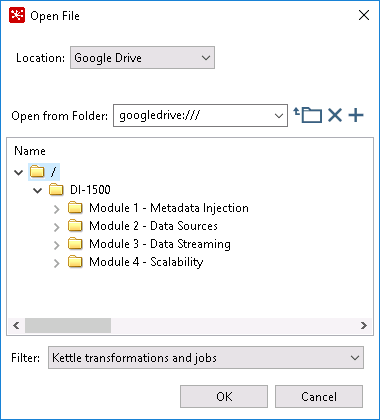
You are prompted to login to your Google account



Once you have logged in, the Google Drive permission screen displays.



1. Click Allow to access your Google Drive Resources.



A new file called StoredCredential will be added to the same folder as the client\_secret.json file. This file will need to be copied to the Pentaho Server plugin credential location.

pentaho-server/pentaho-solutions/system/kettle/plugins/pentaho-googledrive-vfs/credentials