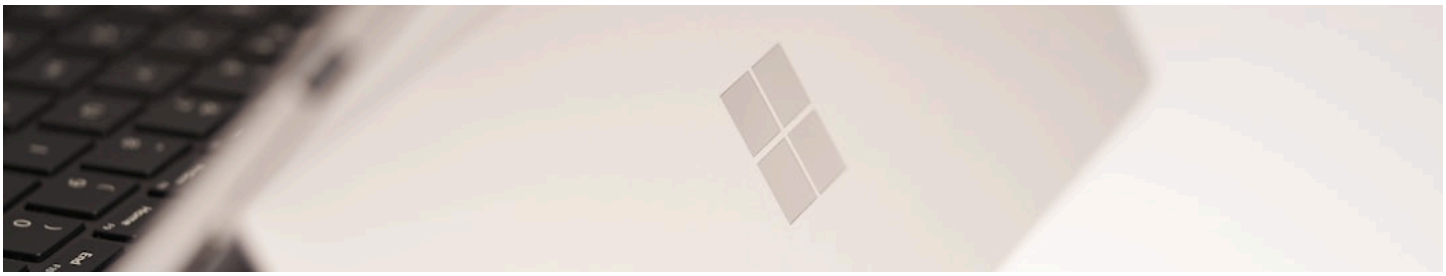


Microsoft File Share Crawler Documentation

Introduction

In this document we will explore how to setup the *Microsoft File Share Crawler* for our Dotnet Crawler. This functionality is not available for our external Java crawlers.

Prerequisites

In order to get the Microsoft File Share Crawler functioning we *must* provide the following, minimum credentials from our windows machine:

- **Username** - the Microsoft Username to access a file-share, prefixed by the domain if applicable (e.g. DOMAIN/BillyBob)
- **Password** - password for this user-account
- **Sever** - the *IP Address or Hostname* of the server hosting the file-share.
- **Share name** - the name of share on this machine.

Username & Password

Provide the relevant username and password required to access the windows account on your machine, this should be sufficient.

Server

For this step, be sure you know how to resolve the machine name (NETBIOS vs DNS) or use the an IP address. You must be on the same network and not blocked by a firewall. You can always try and connect using the credentials from one machine to the target machine (or even on the same machine).

Share-name

The name of the folder you wish to crawl into from your Windows machine. This folder must be shared from the host server using the Microsoft sharing protocol. Once shared, a Windows share looks like:

> *server-or-ip-adddress**share-name*

This is commonly known as a UNC path (Universal Naming Convetion) and is a Microsoft specific notation (not Universal). A file inside a folder on that server then might look like:

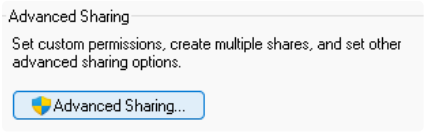
> *server-or-ip-adddress**share-name**folder 1**the file.txt*

NB. a Windows file system will contain Windows 1252 characters. Our crawlers convert such characters to UTF-8 where possible.

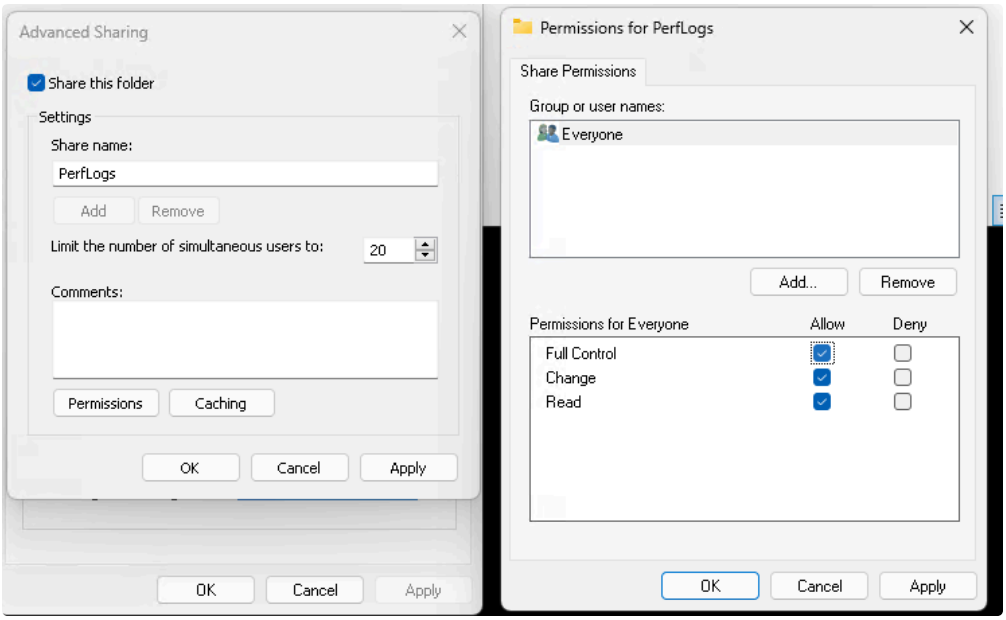
Permissions

Before we move on, we need to be sure that the correct permissions are set for the share folder.

1. Right click on the folder and select *properties*
2. Navigate to the sharing tab, and select *advanced sharing*



3. Tick *share this folder* and then click permissions, check *full control*

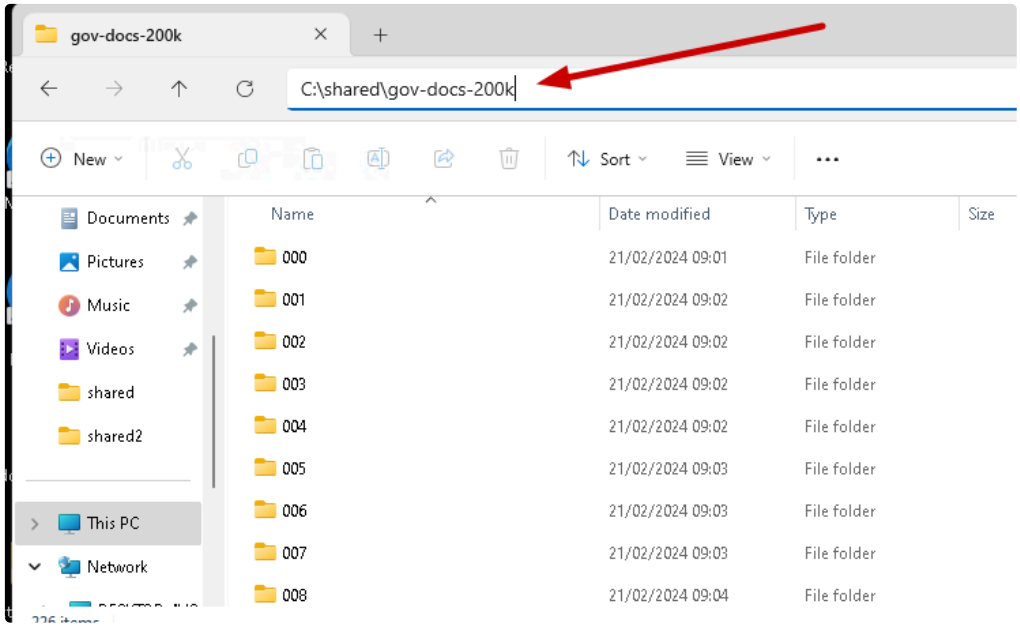


Provide the name of the folder to the crawler.

Optional Settings

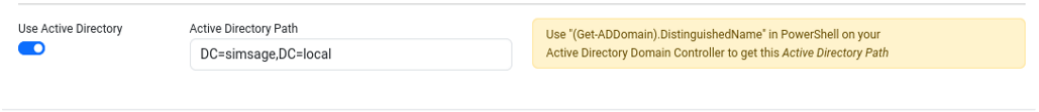
Share Path

By default the crawl *recursively* reads from the root of your share folder, if you wish to crawl from a *specific folder,* you will have to provide the path here, you can do this by navigating to this folder in your file explorer and copying the path.



Using Active Directory

In this section, we will explore the functionality we offer regarding *active directory.*



Use Active Directory enabled

By *enabling the active directory (AD),* you can resolve ACLs. An ordinary Microsoft file system only exposes ACL ids without email addresses. An ACL id can refer to a group or a user.

If enabled, Microsoft Active Directory can be queried. All AD users can read from this directory. By enumerating the folders, user-memberships, and users you can build a list for resolving ACL ids to folder names and user accounts.

It is vital that you set the email addresses for all AD Users on your domain. SimSage can only resolve users by email addresses. The email address is a property of every user in the Domain Users and Groups control panel of Domain Controller.

Here:

- **Active Directory Path** - an LDAP descriptor unique to the domain for reading information from the domain (such as users and folders). This name is set when the domain is created and the usage of DC, OU. This value is mandatory and needs to be correct. It tells LDAP where the server is located. To get this value, Open a PowerShell session on your Microsoft Domain Controller, and run the following command. It will give you the required value.

```
(Get-ADDomain).DistinguishedName
```

  This value should resemble something like “DC=simsage,DC=local” or “DC=myorganisation,OU=com”