



32158 Camino Capistrano
Suite A PMB 373
San Juan Capistrano, CA 92675
Sales@CaptureComponents.com
www.CaptureComponents.com

Sharing Documents on the Internet

With ccScan for Amazon S3

Document capture to the Cloud

Overview

One way to share documents on the Internet is to create a dedicated website and share those documents with the World. Alternatively, the documents can be posted on popular document sharing websites with access to the documents restricted by login names and passwords.

However, there are other ways that do not involve dedicated or shared websites. For example, Capture Components' ccScan, a powerful document capture application for the "Internet Cloud," allows scanning and/or importing of PDF documents into Amazon's S3 Cloud-based data storage platform.

The captured PDF documents can be shared using URLs which are browser links pointing to a specific file on the Internet. These URLs are given an explicit expiration date and time so they will not remain active forever. In addition, since ccScan fully supports the PDF native password mechanism, documents can be captured so that viewing or printing of the resulting PDF files will require a password.

These URLs can be forwarded to interested parties through emails, tweets, embedded in documents, embedded in web pages, and other means. The documents are stored in Amazon S3 (Simple Storage Service) and could be deleted with ccScan once they no longer need to be shared.

Applications

There are many organizations that can benefit from capturing documents with ccScan and sharing those documents easily over the Internet. Although ccScan has impressive document import capabilities (existing PDF files could be converted to password-protected PDF files during import, for example), the most benefit is derived when copies of paper-based documents need to be shared.

Since ccScan is a full featured, production-level document capture system, paper-based documents can be captured efficiently and automatically stored in Amazon S3, all in a single operation.

Applications for this capability are numerous, for example law offices can share case documents with clients and opposing parties and accountants can share tax information with clients.

ccScan For Amazon S3

Capture Components markets different versions of ccScan with each targeted at a different document repository. The ccScan versions for “Internet Cloud” based storage support Google Drive and Docs, Box, Salesforce, Dropbox, Microsoft SkyDrive, Amazon S3, Windows Azure, etc, but other types of document repositories are supported as well. For example, documents can be captured remotely and stored in Microsoft SharePoint and SharePoint Online using SharePoint’s Internet-based web services interface. Additionally ccScan works with the Microsoft Windows file system utilizing its local hard drive or networked folders for storage.

Together Amazon S3 and ccScan form a cost-effective, secure platform for scanning, importing, and storing documents online. In addition, it is easy to view the documents stored on Amazon S3 with ccScan. Since Amazon S3 takes care of redundancy there is no need to worry about backups or disaster planning. The combination of Amazon S3 and ccScan also establishes a high level of security. Documents are transferred between ccScan and Amazon S3 using Secure Sockets Layer (SSL or HTTPS).

Even though ccScan costs significantly less than competing products, it is a full-featured, production-level document capture system. The software takes full advantage of today’s multi-core processors. On a quad-core computer the scanning, barcode detection, text recognition, and export to Amazon S3 each run on a separate core to maximize throughput. Consequently most TWAIN-compatible production scanners are operated at rated speed. Since ccScan has been designed to be enterprise-friendly, it contains a Setup mode and a User mode. The Setup mode is password protected. Jobs are created in Setup mode and can be used over and over again in User mode. To make sure ccScan can be used immediately after installation, without having to create a job first, a set of default jobs is provided to cover common document capture scenarios.

Jobs can be associated with a scanner button, which can make running a job in User mode as simple as pressing a button on the scanner. ccScan also contains an optional indexing/QA mode that allows the operator to check the document or enter indexing data before the document is sent to Amazon S3. Documents are automatically separated if barcode sheets are used.

On platforms that contain no default document viewing options such as Amazon S3, ccScan provides the capability to create View jobs that will list all documents entries or a subset in either a sequential list or hierarchical format. These same View jobs are used also to generate a URL for the selected document for sharing on the Internet.

Remarkably despite its low price ccScan has all the high-end features and no restrictions such as page click charges. For detailed information about ccScan or to download a free trial version, please visit Capture Component’s website at www.capturecomponents.com.

Typical Scenario

To give the reader a flavor of capturing and storing documents in Amazon S3 with ccScan the following sections contain a walk-through annotated with screenshots of a typical scenario. Although the default set of jobs installed with ccScan cover capturing and storing documents in Amazon S3, the example scenario assumes that no ccScan job has been created previously so both Setup and User modes can be shown. For ease of illustration certain details such as the creation of an account with Amazon S3 or how to enter Setup mode have been left out of this descriptive scenario.

Create Scanning Job

To get started a new Scanning job must be created. In ccScan jobs are created in Setup mode. Once this job has been created, it can be used over and over again in regular User mode. The Administrator selects default values for the various job settings and decides which job settings can be entered or modified by the operator. The job settings are organized by category such as scanner settings or export settings.

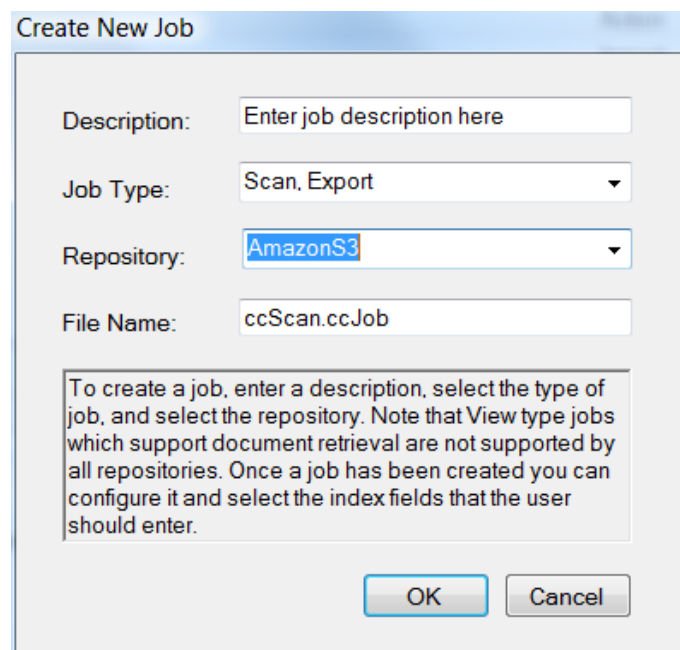


Figure 1 - Create New Scanning Job

In the dialog box above Amazon S3 has been selected as the repository for export of the captured documents. If ccScan was licensed for multiple repositories, the target repository for the job would be selected here from the list of available repositories. Note that the Job Type is Scan, Export.

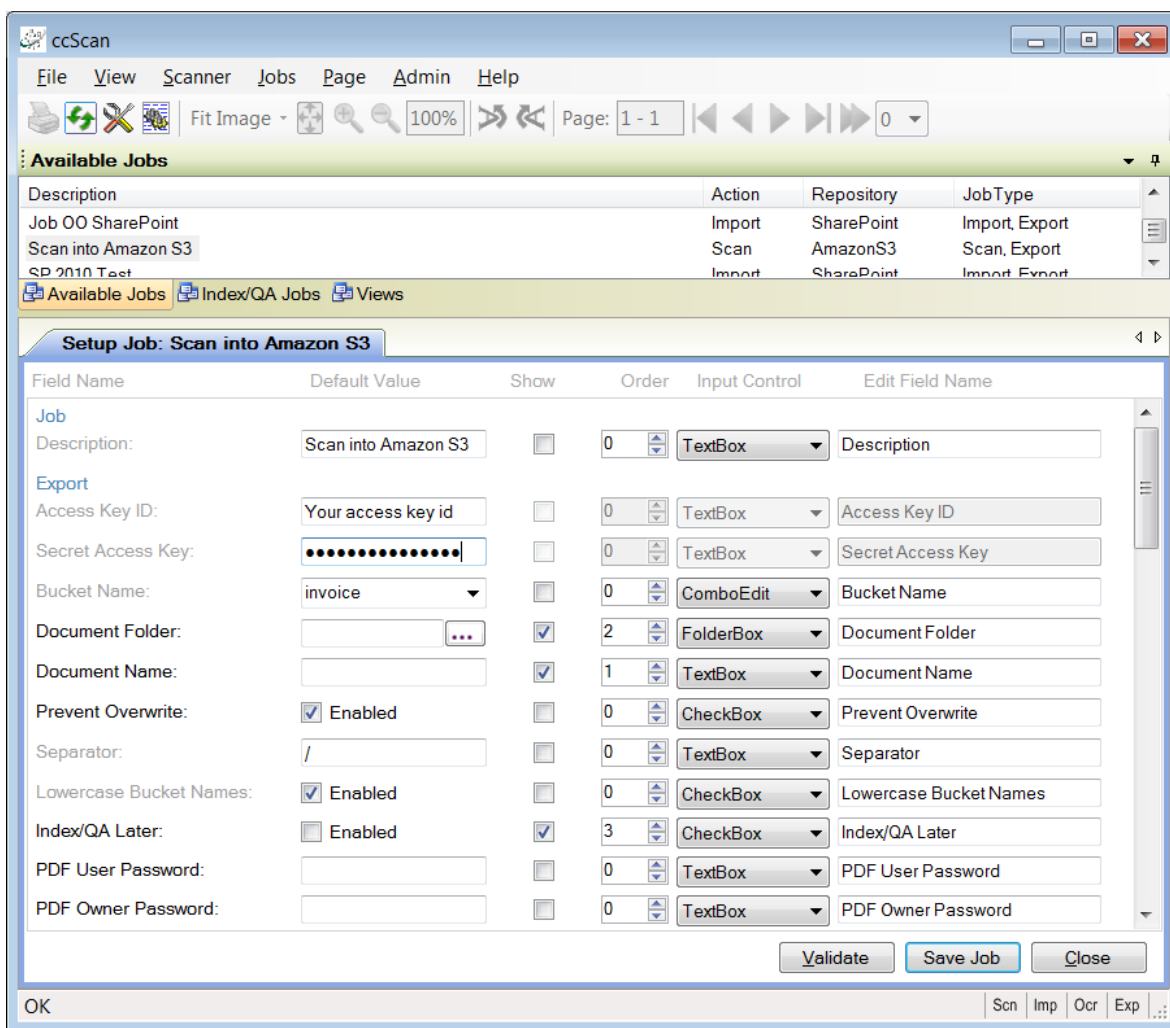


Figure 2 - Configure New Scanning Job

In the form above the Administrator can configure the new Scanning job. The same form is used to make modifications to the job later. In the Export Settings, enter the first Amazon S3 account information, such as bucket name, access key, and secret access key. Note that normally the Amazon S3 account information is entered during first time startup when the default set of jobs is configured for immediate use.

The name of the newly scanned document is entered in Document Name. Optionally, a folder name can be entered in Document Folder. To allow for automated naming ccScan supports the use of “macros” in document and folder names that automatically insert counter and barcode values. Automated naming is used with high-capacity scanners whose hoppers can hold large stacks of documents separated by barcode sheets. Prevent Overwrite is another Export Setting which if checked will report an error if an existing document is about to be overwritten. Also provided are a series of standard PDF settings, such as user and owner password, which will be applied to the captured documents before they are exported to Amazon S3.

For each setting, the Administrator selects or enters a value. Default values are provided where possible. The Administrator can also make settings available for the operator to enter, except for the Administrator-only settings. For the settings to be entered by the operator, the Administrator determines the order in which these settings are shown, optionally modifies the default description of each setting, and selects what type of input control should be presented to the operator.

After the Export Settings, the Scanner Settings are entered. The scanning of documents in ccScan is based on the TWAIN standard. The scanner manufacturer provides TWAIN drivers. All the scanner settings you would expect in a production-level document capture system are provided: basic settings such export format, resolution, paper size, and advanced settings such as barcode detection, which for example, can be used to automatically separate documents or automatically specify folder names. Jobs can also be associated with a button at the scanner front panel so that the job is started automatically when the button is pressed.

TWAIN settings highly specific to the type of scanner can optionally be set with the TWAIN dialog box. These settings are saved locally on the computer and associated with the job. For those settings common between ccScan and the TWAIN dialog box, the ccScan settings can be given priority over the TWAIN settings. With this capability the administrator can give the operator access to the TWAIN dialog box and still for example enforce company standards such as 200 DPI resolution.

Run this Job

Once the job has been created it can be run over and over again in User mode. The job is started either from the scanner front panel or with the Start Job button in ccScan.

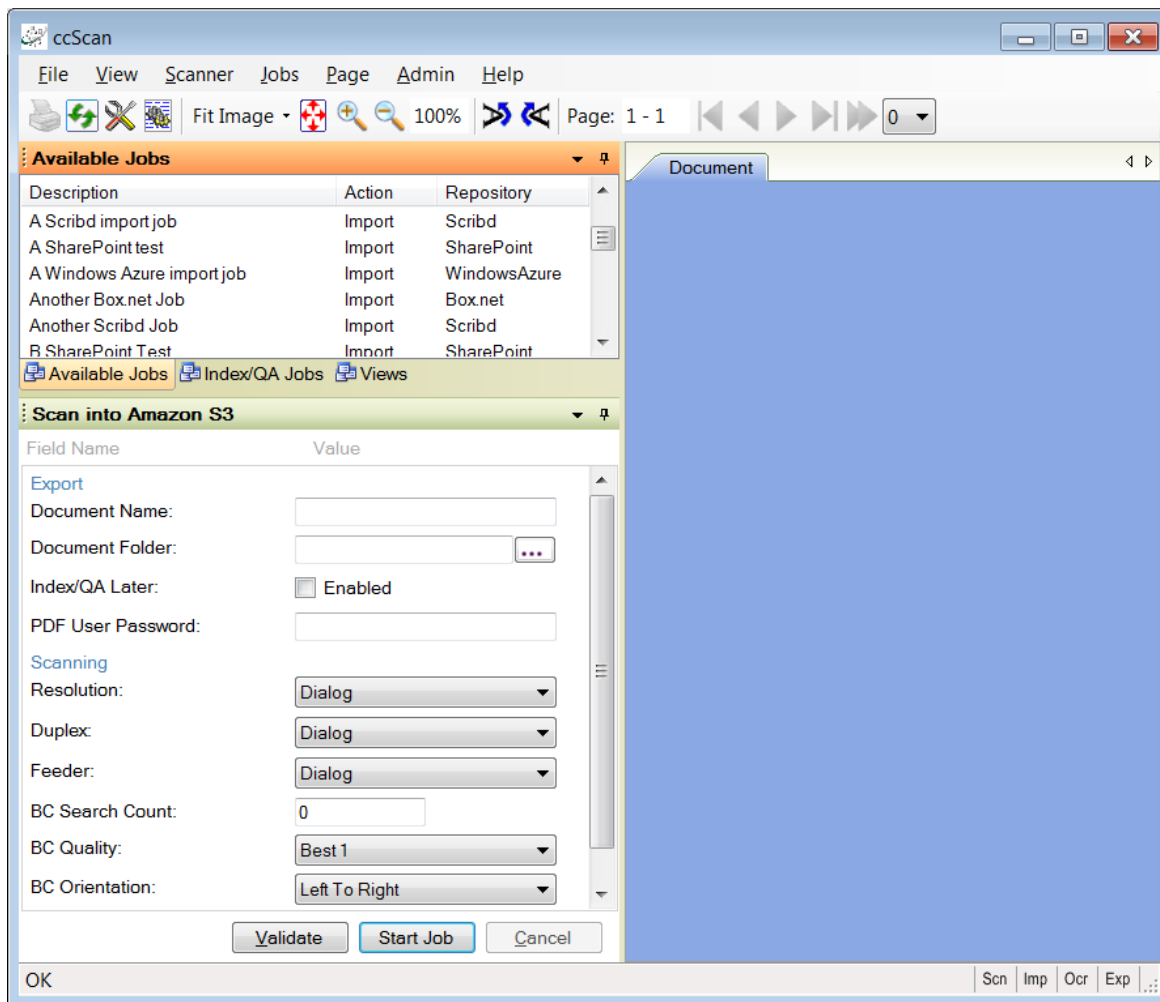


Figure 3 - Run a Job

Typically jobs are configured so that the operator must enter at least a document name, unless ccScan’s automated naming capabilities are used. Optionally the operator can change default scanner settings or enter an Amazon S3 folder name where the captured document will be stored.

Once the job is started, ccScan disables the Start button and does not allow any changes to the form. However, the operator is allowed to press the Cancel button. To provide visual feedback the scanned pages are displayed while the job is running.

Create a View Job

A new job must be created to view the captured documents in Amazon S3.

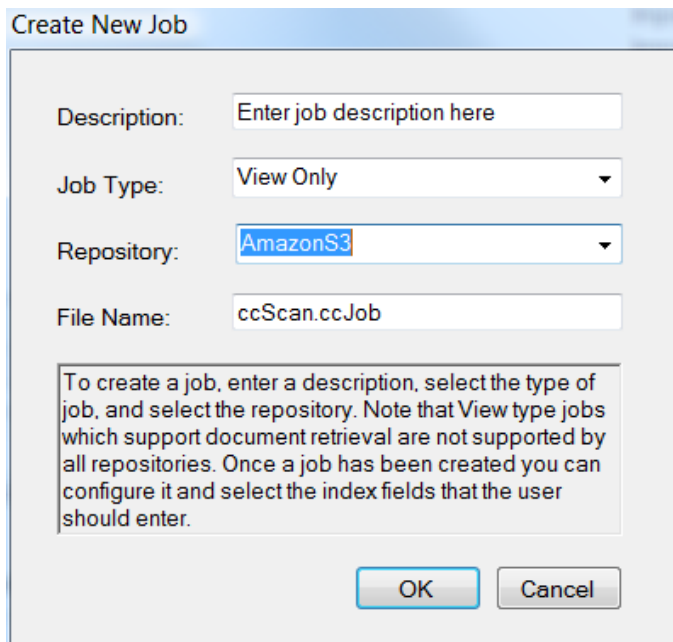


Figure 4 - Create New View Job

In the dialog box above Amazon S3 has been selected again as the repository. Note that the job Type this time is View Only.

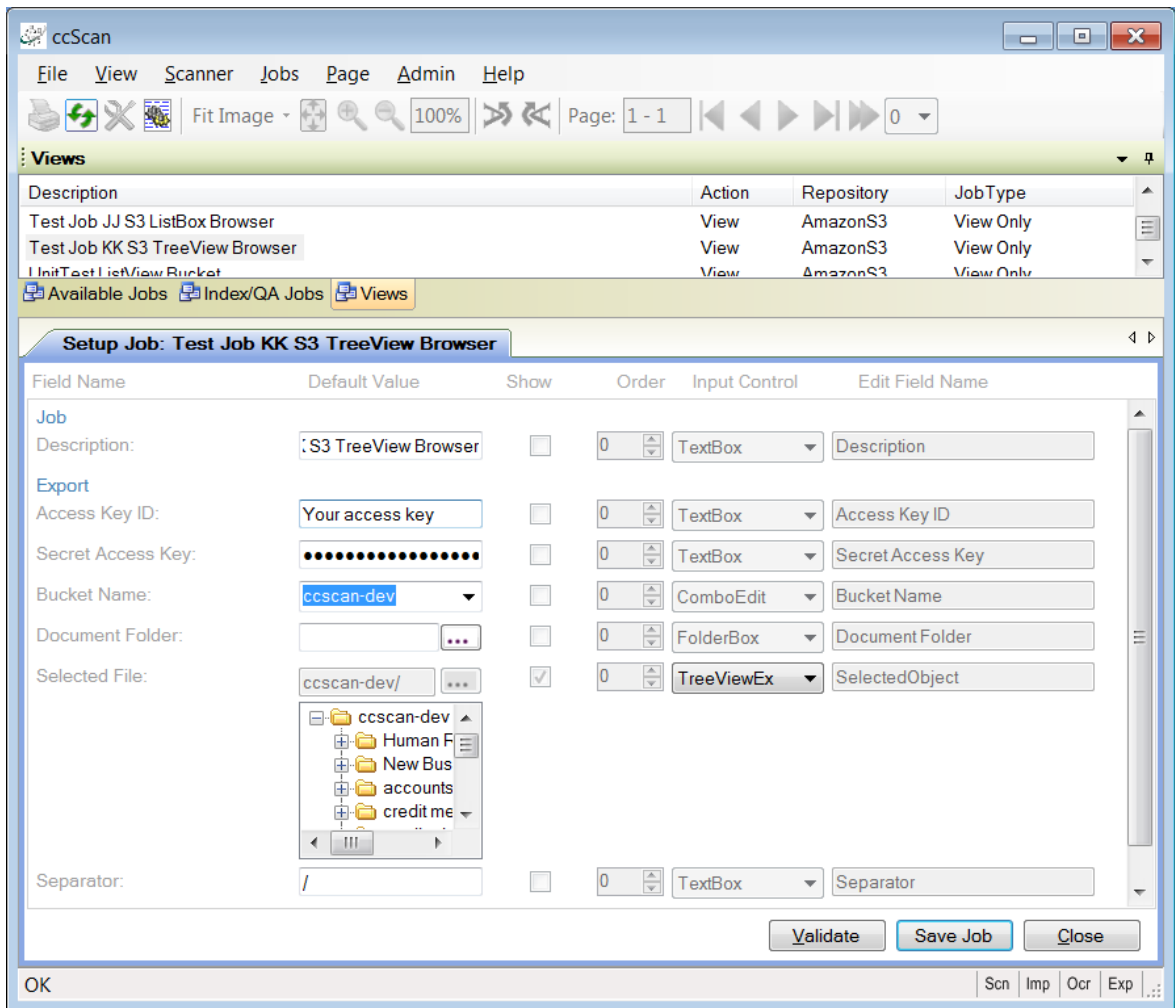


Figure 5 - Configure New View Job

In the form above the Administrator can configure the new View job. The same form is used to make modifications to the job later. In the Export Settings the same Amazon S3 account information such as bucket name, access key, and secret access key are entered. Other than the document folder, which is used to restrict the view to only the documents in this folder, there is not much to configure. The Administrator can choose between a list view control and a hierarchical tree view control. The extended versions of these view controls allow the operator to navigate to different folders in Amazon S3.

Select a View Job

Once a View job has been created it can be selected in User mode. View jobs, unlike Scanning and Import jobs are not really “run.” Instead ccScan lists those jobs in a separate tab, called Views in the Jobs window. Just selecting the job will cause the list view or tree view control in the form window to be populated with the names and attributes of the captured documents in Amazon S3. To avoid having to retrieve the file list from Amazon S3 over and over again, which could take some time if a large number of documents have been captured, ccScan employs a sophisticated caching mechanism. The cache maintains a list of documents in the computer’s memory for quick access.

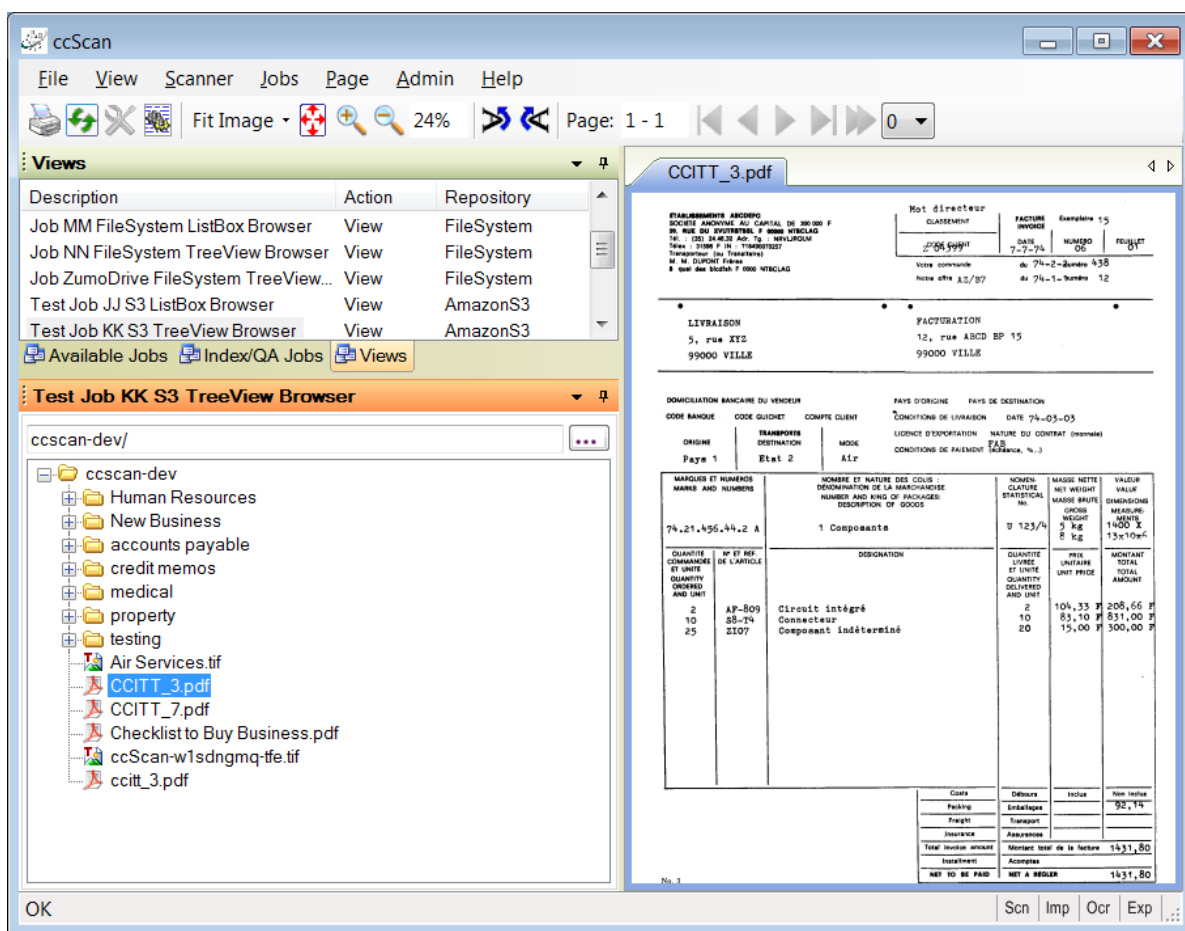
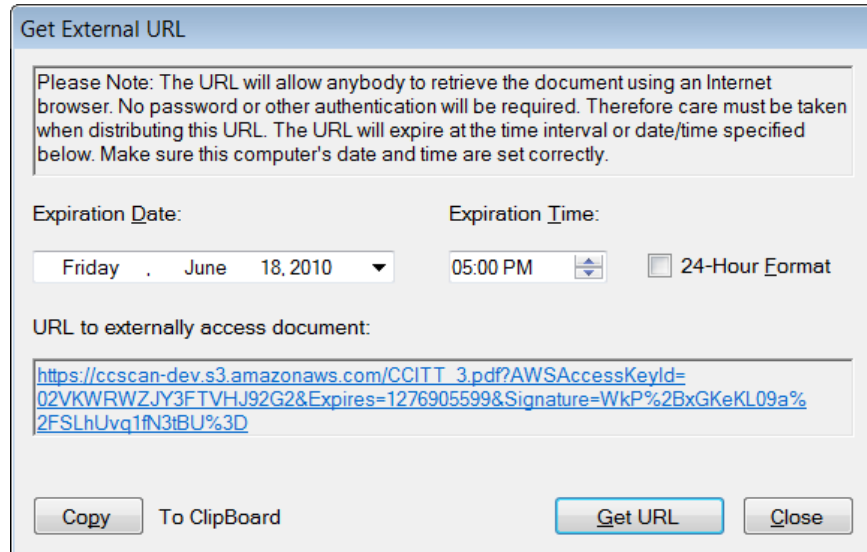


Figure 6 - Selecting a View Job

When multiple View jobs are present the operator can easily switch between them by just selecting the job of interest.

Create a Document URL

Once the previously captured documents are listed they can be viewed, printed and deleted by selecting them and either using the File menu or right-click context menu to select the desired action. For documents in Amazon S3, it is also possible to create a URL for the selected document.



Get External URL

Please Note: The URL will allow anybody to retrieve the document using an Internet browser. No password or other authentication will be required. Therefore care must be taken when distributing this URL. The URL will expire at the time interval or date/time specified below. Make sure this computer's date and time are set correctly.

Expiration Date: Friday, June 18, 2010

Expiration Time: 05:00 PM 24-Hour Format

URL to externally access document:

https://ccscan-dev.s3.amazonaws.com/CCITT_3.pdf?AWSAccessKeyId=02VKWRWZJY3FTVHJ92G2&Expires=1276905599&Signature=WkP%2BxGKeKL09a%2FSLhUvq1fN3fBU%3D

Copy To Clipboard Get URL Close

Figure 7 - Create URL

The URL must be given an explicit expiration date. Once created, the resulting URL can be copied to the clipboard and pasted in emails, tweets, web pages, and other sources. Since ccScan fully supports the native PDF password mechanism, documents published in this manner can be captured so that they require a password for viewing and printing.

Document URL in Browser

This screenshot below shows the Internet Explorer displaying a URL created with ccScan.

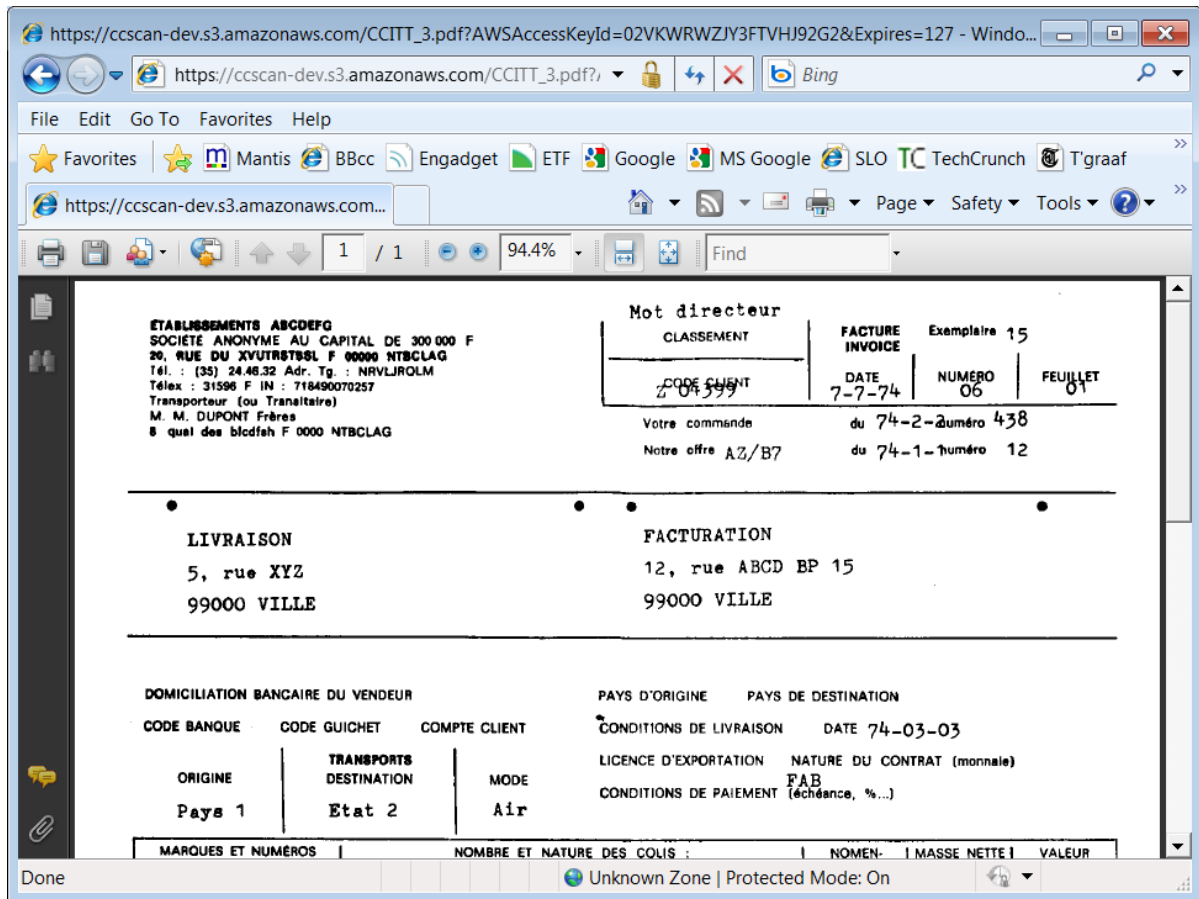


Figure 8 - Internet Explorer

Annotation and Redaction

It is not uncommon to annotate or redact selected paragraphs of documents that are shared. For example, redaction is used routinely in the legal sector. To allow review, correction, or annotation of documents before they are stored in Amazon S3, ccScan provides an optional Index/QA mode. This mode can be enabled on a per-job basis by checking the Index/QA Later field. When enabled the documents produced with a Scanning or Import job are cached locally and made available in the Index/QA tab of ccScan.

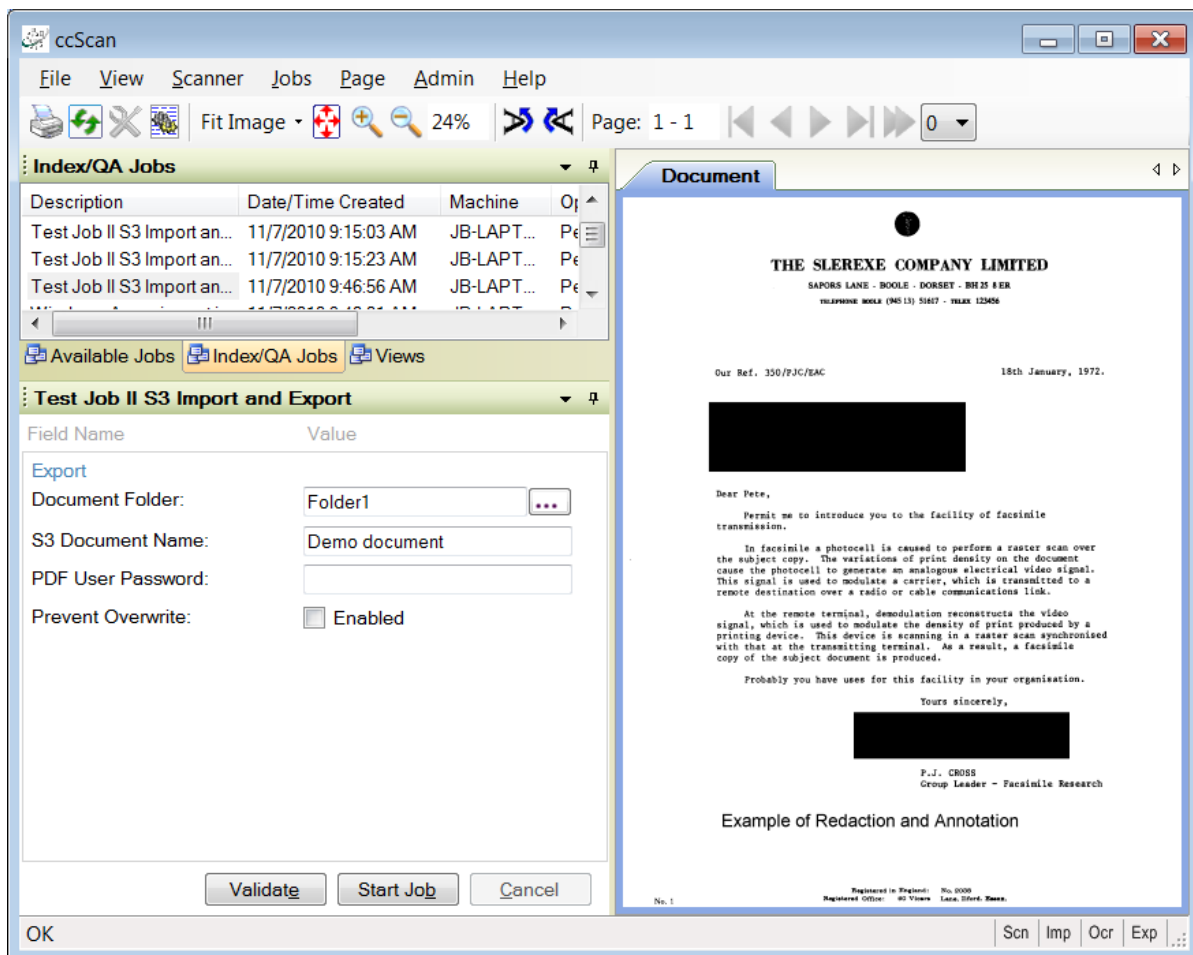


Figure 9 - Index/QA Mode

Before releasing the documents to Amazon S3 from Index/QA, descriptive information can be added to index fields, and pages can be deleted or inserted. Index information can be entered manually using the keyboard or extracted automatically from selected zones of the image using OCR (Optical Character Recognition).

In addition, pages can be annotated with text and selected portions of the pages could be redacted. Note that the annotations and redactions become permanent once they are saved and cannot be removed later.

Afterword

We at Capture Components LLC hope that this White Paper is of benefit to you. If it is and you are able to easily capture and share your documents on the Internet, then we have achieved a measure of success. If you choose to use ccScan as your document capture application, we will do our best to provide you with the highest quality product and support.

Visit us at <http://www.capturecomponents.com>