PSI:Capture – Strategies for Spreading Auto Import Processes

Overview

The Auto Import service in PSI:Capture allows for automated processing of documents. The Auto Import service monitors locations for incoming files, which are then collected into queue directories for processing. When a queue directory is created, a reference to the directory is stored internally in PSI:Capture. The Auto Import service works through this list of queue directories one at a time, generating a new batch for each, and then importing and processing the contents of the directory.

PSI:Capture is installed in one of three ways: in a **stand-alone** mode, in a **server** mode, or in a **networked** mode working in concert with a PSI:Capture server. Stand-alone and server installations handle the duties of monitoring, queuing files, and maintaining the list of queue directory references. They can also create and process the resulting batches.

Machines with network installations can be called upon by the server to create and process batches as well when the server is busy, as shown in Figure 1 below. The more network stations you have configured for Auto Import, the more batches your environment can process simultaneously.
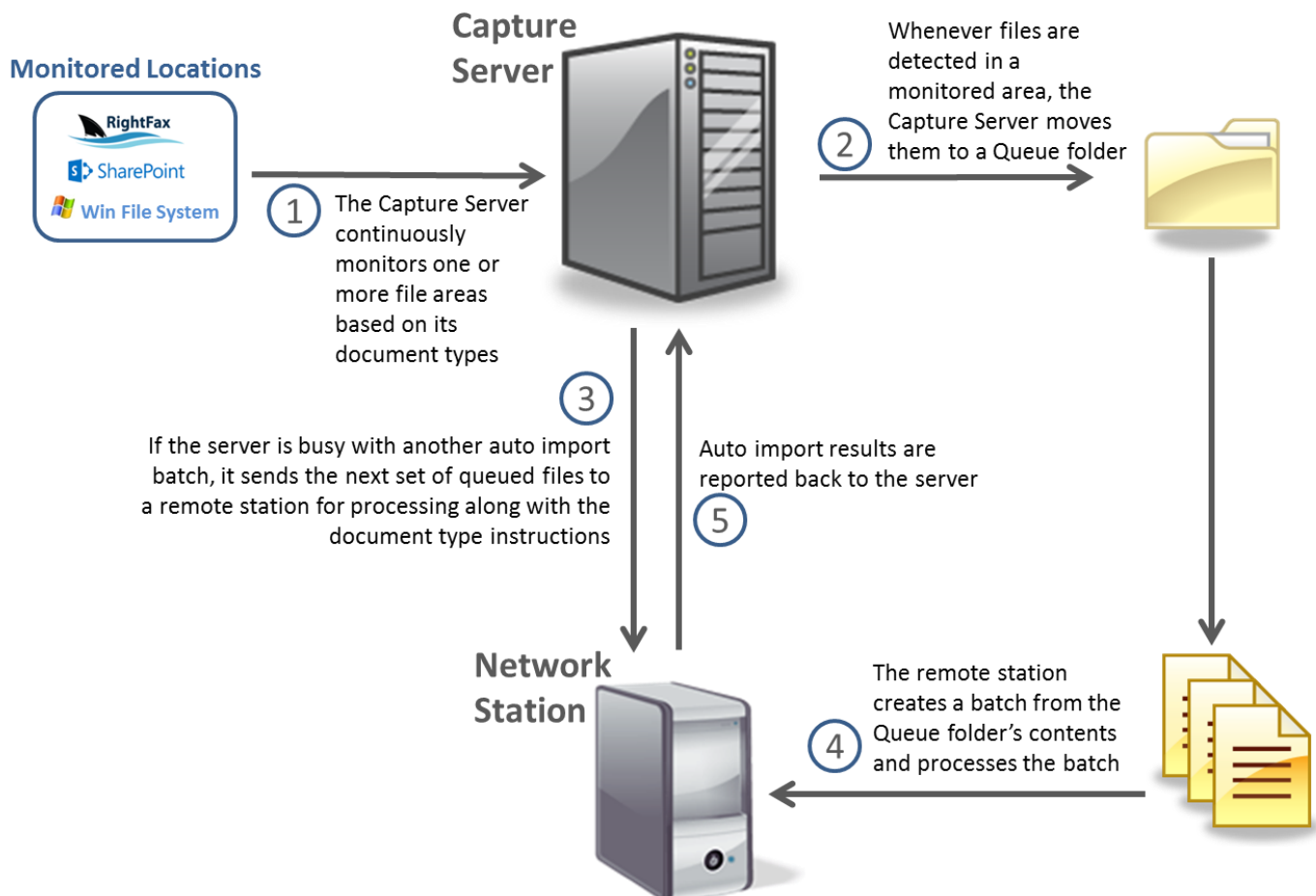

Figure 1 – Classic Auto Import with a remote workstation

PSI:Capture – Strategies for Spreading Auto Import Processes

Spreading the Auto Import Load

In most high-volume situations, use a PSI:Capture server with PSI:Capture network installations on other machines as depicted in Figure 1. The server contains the master list of document types, and handles all monitoring and queuing duties. Once a queue is ready to become a batch, the server can either create and process the batch or have one of the networked machines handle the task. In this way multiple queues can be processed simultaneously.

Although less common, it is also possible to use multiple PSI:Capture servers or stand-alone installations to spread auto import loads (Figure 2). **Each must have its own set of document types and separate monitored areas – multiple machines must never monitor the same area.** The maintenance difficulties inherent with such an arrangement make this a less desirable option.

Because a stand-alone PSI:Capture installation cannot interact with other PSI:Capture machines, it is unable to share its auto import load with other installations.
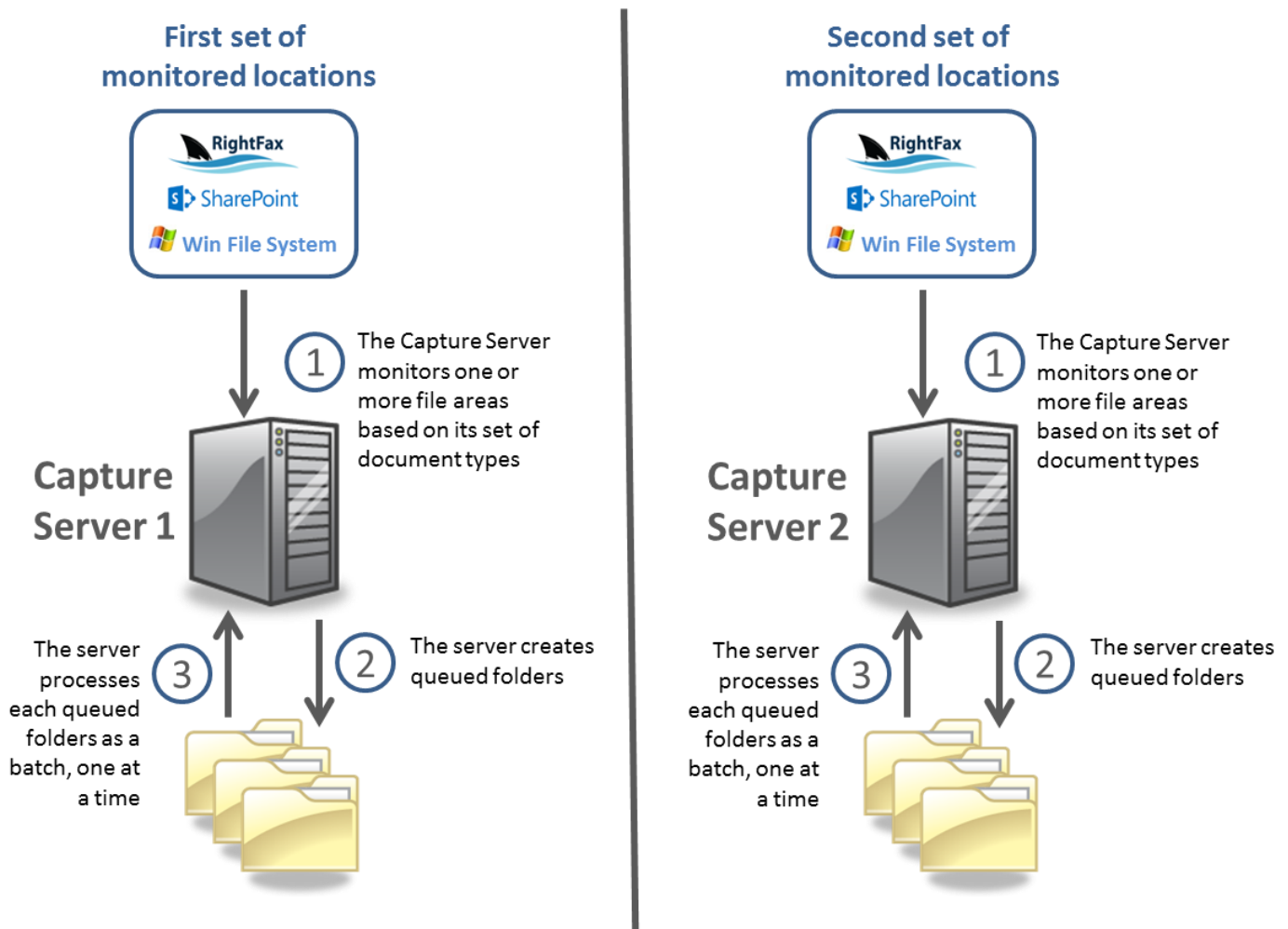

Figure 2 – Load sharing among multiple PSI:Capture servers

PSI:Capture – Strategies for Spreading Auto Import Processes

Workflow Auto Processing

PSI:Capture version 5.0 introduced an automation enhancement called Workflow Auto Processing (WAP). When enabled, each PSI:Capture workstation can be tailored to automatically process workflow steps for existing, checked-in batches in the central Batch Manager. This allows the PSI:Capture Auto Import service to detect and complete these batches without additional human intervention.
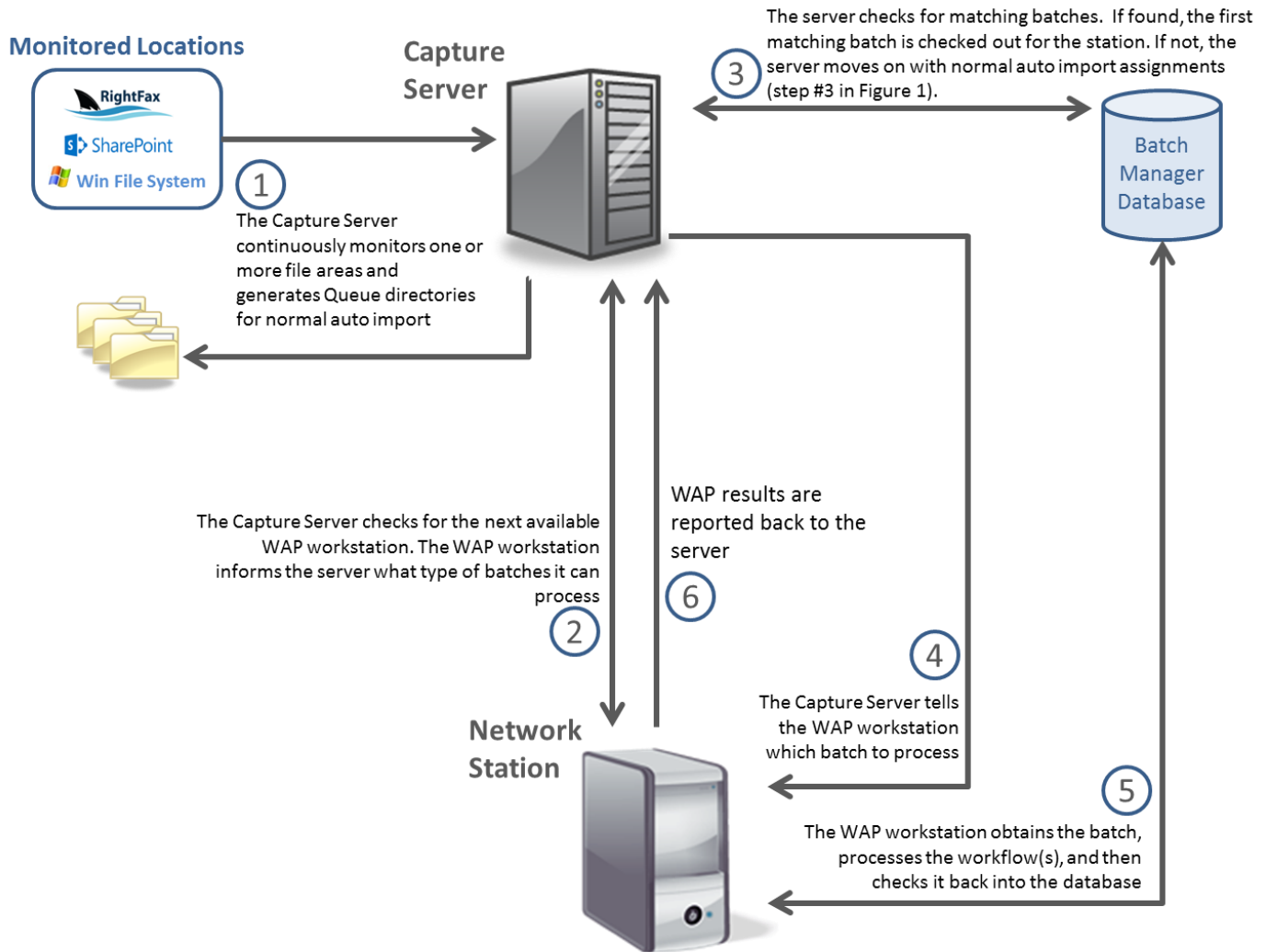

Figure 3 – Details of Workflow Auto Processing

Machines are configured for WAP in one of two modes. The first completes batches whose document type is set to auto-process through the entire workflow (*Attempt to Auto Process Batch through Workflow* checkbox on Step 9 of 10). If a batch stops processing due to an exception or reaches a manual step (such as Quality Assurance), this mode is designed to detect when such batches are ready to proceed with the next workflow step, and then resubmits them for auto-processing.

PSI:Capture – Strategies for Spreading Auto Import Processes

The second mode works with batches whose document types are not set to auto-process batches through the entire workflow. Each machine in a server / network configuration can be configured to complete one or more workflow steps for these batches. For example, machine #1 can be configured to process OCR workflow steps, while machine #2 can be reserved for Indexing steps.

Machines can be configured exclusively for WAP, exclusively for auto import, or for both WAP and auto import. When configured for both WAP and auto import, WAP takes precedence.

Load Sharing Considerations with Workflow Auto Processing

Workflow Auto Processing (WAP) can have either a positive or a negative effect on overall load sharing. A machine can either process a queued set of files for import or process a batch through Workflow Auto Processing; it can't do both simultaneously.

Because WAP takes precedence over import, having WAP enabled on every machine in your networked environment has the potential to slow down the normal auto import process. If you have hundreds of batches waiting for completion in the Batch Manager, the available auto-import workers will be consumed with WAP duties. This leaves no workers available for importing files. New files will continue to queue, but none of the queue directories will be processed into batches until workers become available.

Having too few WAP stations can have a similar affect on overall batch completion. Files in queue directories will be quickly converted to batches. However, incomplete batches will build up in the Batch Manager, requiring more human intervention and thus longer completion times.

When using Workflow Auto Processing, make sure you have the proper mix of auto import and WAP workstations in your environment to ensure smooth operations.

PSI:Capture – Strategies for Spreading Auto Import Processes

The Auto Import Service

The diagram below provides an overall view of the entire Auto Import Service.

- The PSI:Capture server's Auto Import Service handles all duties related to directory monitoring and file queuing.
- Network stations can import files or process batches assigned by the PSI:Capture server.
- Each network station that performs imports must have full access to the monitored areas where queued directories are stored.
- Each network station registers itself with the PSI:Capture server's **Networking Monitor** at startup. The Networking Monitor maintains a **Worker List** of available network stations that can handle extra Auto Import and/or Workflow Auto Processing duties for the server.
- The **Directory Monitor** watches for incoming files in the monitored directories defined in the PSI:Capture server's document types. When new files are detected, the files are moved to a new queue folder. A reference to this queue folder is stored in the **Queued Folders List**.
- The **Workflow Auto Processing/Queue Monitor** is in charge of assigning Workflow Auto Processing and Auto Import work to the next available worker in the Worker List.
- Once a worker completes its assigned task, it reports its results back to the server and is made available again for additional work.