



White Paper: Clustering of Servers in ABBYY FlexiCapture

By: Jim Hill
Published: May 2018



Introduction

Configuring an ABBYY FlexiCapture Distributed system in a cluster using Microsoft Windows Server Clustering provides many advantages. Each server (called node) in the cluster can be configured either for failover or network load balancing. The distributed version of FlexiCapture has been architected to take maximum advantage of these features. There are licensing implications when implementing clustering and/or load balancing, so please contact your ABBYY account manager for details.

Key Advantages of Configuring Cluster in the ABBYY FlexiCapture System

There are several advantages of clustering the ABBYY FlexiCapture Distributed application. The primary benefit of clustering is the provision of fault tolerance and distributed workloads as discussed below.

1. **Fault Tolerance.** Configuring FlexiCapture in a cluster provides greater fault tolerance as each function is distributed between multiple servers making the whole installation much more fault tolerant. If one server is unable to provide a particular function that can automatically be picked up by the other server in the cluster assigned to that function. This eliminates a single point of failure for the system and provides the high availability necessary for a global capture application.
2. **Distribution of Workloads Among Servers.** Configuring FlexiCapture in a cluster provides a very easy way to increase processing capacity therefore making scaling of processing capacity very easy. Just add another server node to the cluster.
3. **Greater Availability.** This is important for companies where FlexiCapture is needed on a 24 x 7 basis such as when the system is used for the global capture application or in cases where critical business functions are provided such as invoice order processing.
4. **Easier Management and Administration of the System.** Administration of the cluster is done centrally through the Microsoft Failover Clustering utility and the NLB Manager in Windows Server. Configuration of clustering on the FlexiCapture system is easily performed through out of the box functions. Additional FlexiCapture licenses may not be required, this is discussed later in this article. But please check with your ABBYY partner or sales representative to be certain before planning the implementation.

Understand Cluster Types

Two Cluster Types and Multiple Configurations

NLB Versus Failover. Microsoft provides two distinct types of cluster configurations, the NLB (network load balancing) type and the failover cluster. The NLB type is used to progressively scale the FlexiCapture system as increasing processing needs are addressed. The failover cluster type picks up the processing task when an operation fails on a single node in the cluster. Note that failover and NLB clusters cannot work on the same server installation.

Active Versus Passive. Each cluster type can be configured in either an active or passive type. In the active failover cluster configuration, each node (server in the cluster) is running and performing processing functions at the same time. In the case of a passive failover cluster configuration the second or successive nodes don't come into action until the previous node fails. In the case of FlexiCapture the failover cluster is generally configured in passive mode for things like the licensing server. More will be explained about this later in this article.

Hardware Considerations

Anyone who has ever worked with a Microsoft cluster will remember that two network cards are required on the server in order to provide a distinct network connection for each node because one is required to handle the private network traffic on the node with two-way heartbeat information.

Basic Configuration of Servers and Nodes

Figure 1 shows the various FlexiCapture stations and servers in the FlexiCapture Distributed system. For high capacity processing each function must be installed on a separate server as shown in Figure 1 below. Keep in mind the distinct differences between the two types of clusters, the network load balancing (NLB) and the failover cluster.

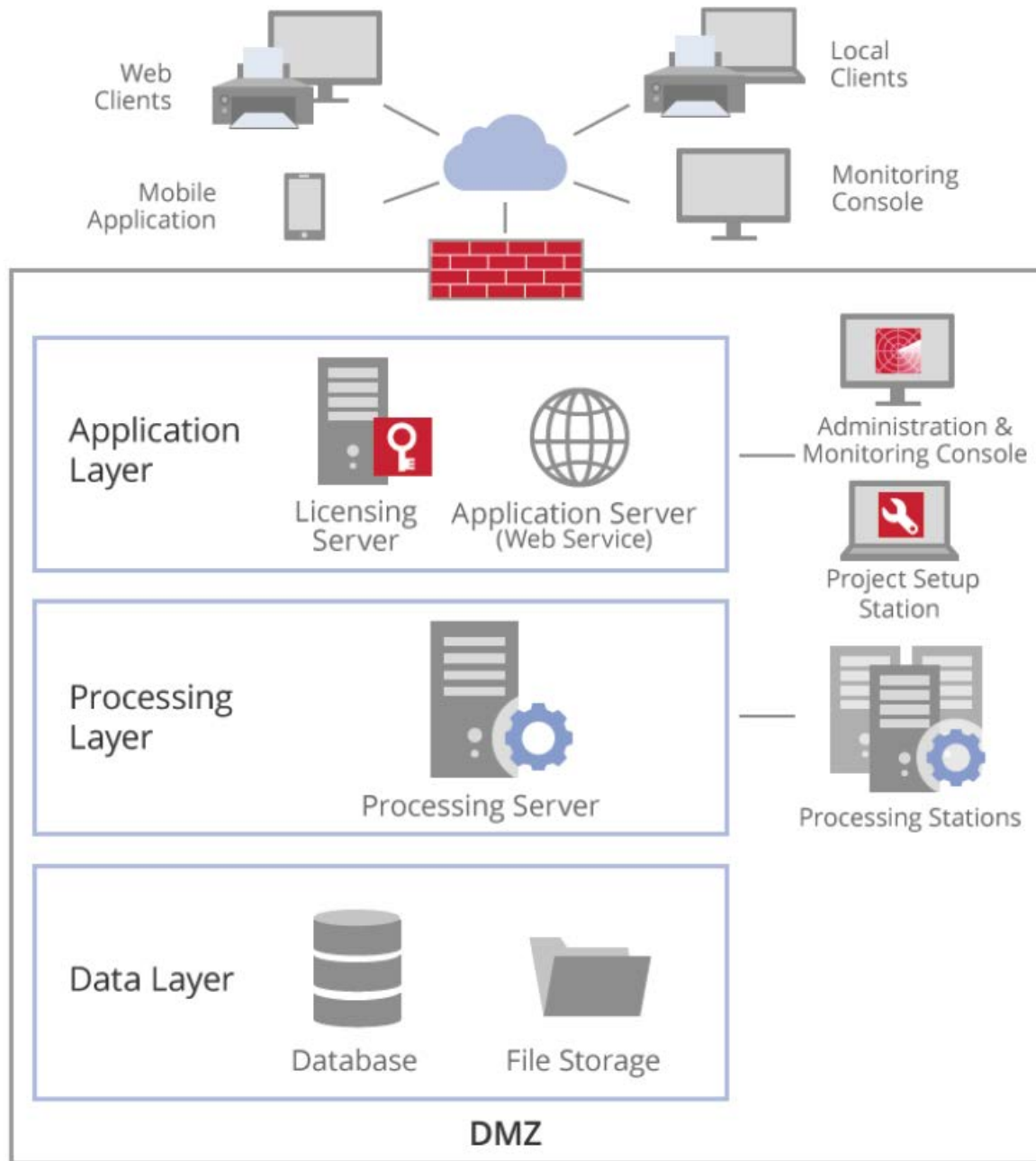


Figure 1: Basic FlexiCapture Station Diagram

Application Layer

- **Application Server**, can be installed on a **network load balancing** cluster. Provides a web service in IIS (Internet Information Services) that verifies user authentication and authorization and performs other functions including execution of the capture workflows. This server hosts

the admin and monitoring console web page through which the main functions of the FlexiCapture system are administered. As an alternative configuration, it is possible to configure the routing of network connections to specific cluster nodes in contrast to using NLB affinity settings. When you first install the application server, you will of course have to create the database and also configure the file storage for this application server. This is done through the web administration console familiar to anyone who has ever installed FlexiCapture Distributed, and hopefully you have provided sufficient DBA credentials for the database creation! Note: scripts are provided in the installation directory (C:\inetpub\wwwroot\FlexiCaptureXX\Server by default) that can be given to your DBA for manual database creation if security is a concern.

- **Licensing Server** (also called Protection Server), can be installed on a **failover cluster**. Manages the software application licensing for all users including those logging in through the various rich clients and web stations. Note that each licensing server must have its own FlexiCapture license (or the same production license with an additional activation) because if one node fails the second node will require a license.

Processing Layer

- **Processing Server**, can be installed on a **failover cluster**. This manages a pool of processing stations which provide the software OCR and other functions. The processing server is actually the only service which is shared between the nodes which makes it easy to test the failover service.
- **Database Server**. FlexiCapture requires access to a Microsoft SQL Server or Oracle instance, except that for development instances Microsoft SQL Server Express can be utilized. Also, for development it is possible to install all servers and stations on a single machine with the understanding that processing performance will be limited. Clustering of the database server is outside of the scope of this paper as it falls into the category of database administration. However, FlexiCapture can work with SQL Server installed on a **failover cluster**. ABBYY recommends that the database server always be installed on a separate machine.
- **Processing Stations**, can be installed on an **NLB cluster** and The web-based stations (which connect to IIS) can be installed on **NLB clusters**. Processing stations are the workers which provide the necessary OCR processing power in order to accomplish the data extraction functions within the allotted timeframe. There is much to understand about choosing the number of cores and memory assigned to processing stations. For example, the addition of CPU cores will only allow for the simultaneous processing of *additional* batches, not faster processing of existing batches. There is a point of diminishing returns when adding cores if the underlying system cannot support them. Almost more important than adding cores is disk bandwidth, because at some point the cores will be limited by the bandwidth of the disks.

It is a best practice to install the processing stations on their own servers. Multiple processing stations on multiple servers inherently provides both load balancing and fail over. Although stations

can also be limited to specific tasks. For instance, you can have a processing station server that only does import, 3 that only do recognition, one for export, etc.

Data Layer

- **File Storage Server.** FlexiCapture requires a network location to store files during the work in process. For a development system or smaller FlexiCapture processing environments this function may be performed by storing files within the database server. Files are only stored in the file storage server as long as steps are remaining to be completed in the workflow. For example, verification operators in an invoice processing solution may need to gather required information about a particular invoice document for a week or longer. During that time the invoice document will remain within the FlexiCapture file storage server. The moment that the last remaining workflow step has been completed the document and data is exported data will remain in the system until the set retention window is exceeded. The default retention window is 14 days, and it is possible to configure the system such that data is never deleted. For extremely high processing environments it is recommended that an external storage server (NAS, the lower cost option, or SAN) be utilized which provides read/write access at 1 Gb/second. In a medium capacity environment, a disk array in a RAID10 (or less ideal RAID1) should be provided along with very fast disk drives or SSD units.

Figure 2 shows a high-level diagram of the basic servers and cluster types.

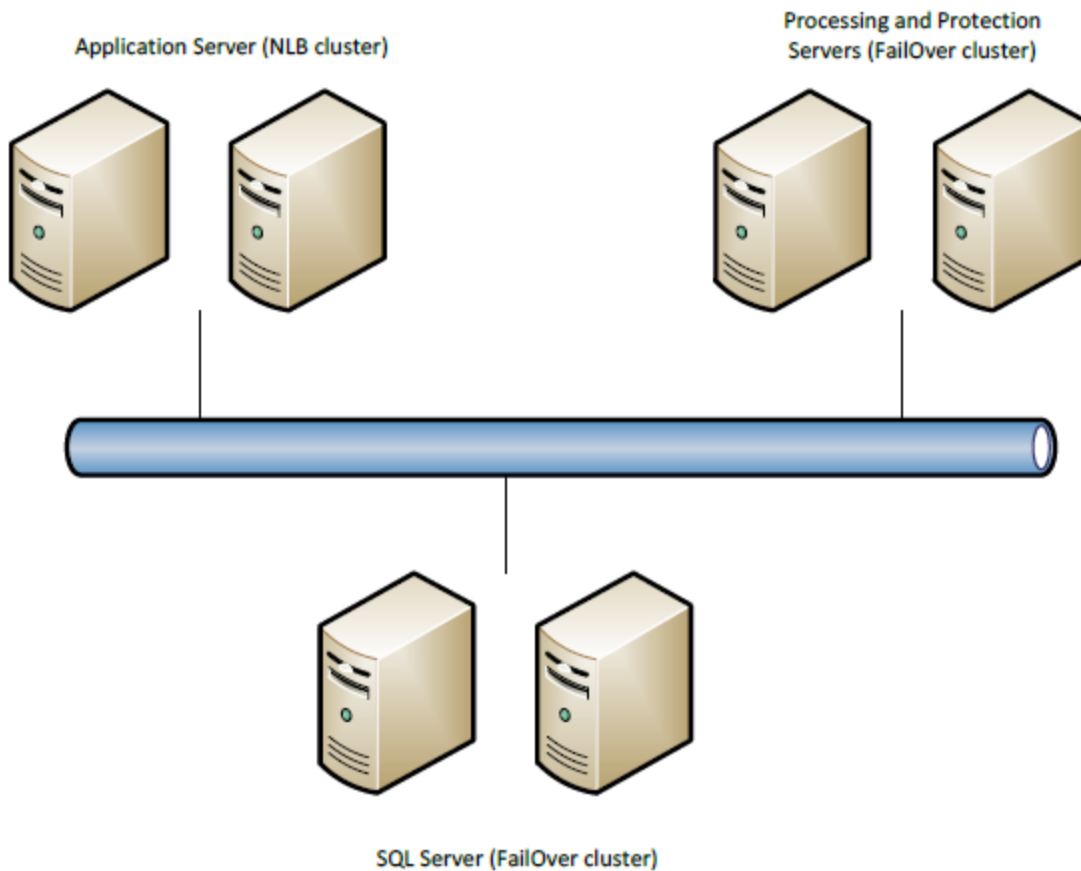


Figure 2: Servers and Cluster Types

ABBYY FlexiCapture Licensing Implications

Licensing is an important consideration in the planning of the cluster. As mentioned in the introduction, there are some licensing implications to cluster ABBYY FlexiCapture. When ordering a cluster license, you must consider both the page count and the licensing of the various stations. This is because during the time that the primary node is unable to serve the licensing function the secondary node will be required to provide sufficient licensing for the users. Considerations include the number of pages expected, the type of FlexiCapture system (invoices versus plain FlexiCapture), and the number of type of each station license required.

The following diagram is useful to help plan the licensing needs for the failover server operating the licensing function. You will want to create your own detailed map of the servers being clustered with details including the main IP address, heartbeat IP address, and DNS assigned server names for each.

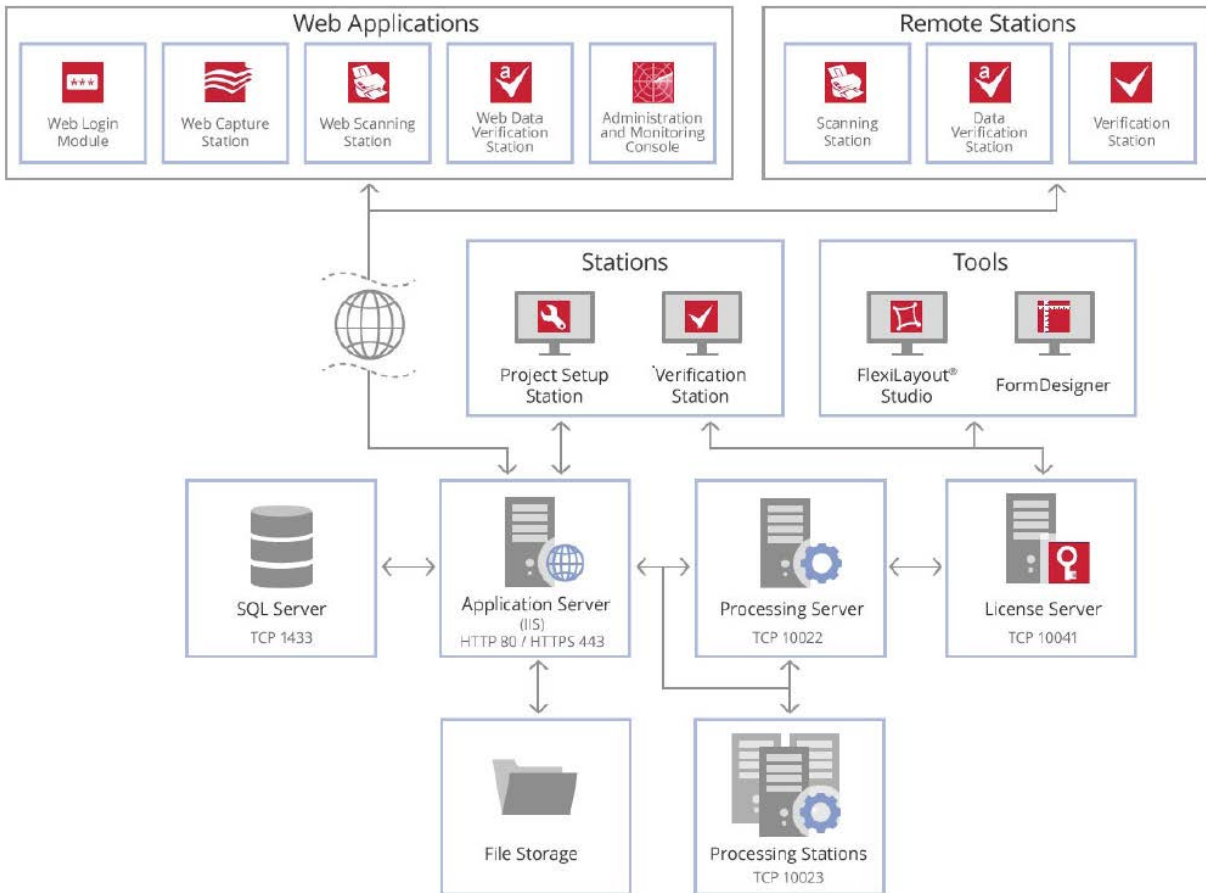


Figure 3: Architecture Showing Stations

Conclusion

Configuring clustering of the servers used in ABBYY FlexiCapture provides many advantages including high availability, load distribution among the servers, easier management and centralized administration. Clustering is a built-in function for FlexiCapture and it can be done for any of the servers without additional licensing costs except for the licensing server itself.

How User Friendly Consulting Can Help

We provide consulting services for ABBYY FlexiCapture, including both the installation and administration of the product well as assisting with or performing development using the web service API. Please reach out to us if you would like to have us solve your document conversion or data extraction challenge or just for help with properly configuring your existing ABBYY FlexiCapture system. We provide a wide range of consulting options including ABBYY trained and certified personnel as well as a wide range of training options to get your employees up to speed on the product very quickly. We

also distribute other ABBYY products such as ABBYY Recognition Server that are geared towards bulk document conversion or language translation.

