Laserfiche Hardware Planning and Specifications

White Paper

September 2012

Laserfiche®

Table of Contents

Introduction

This paper covers how to effectively gather requirements, survey existing resources, and plan system and hardware needs for a Laserfiche installation or integration. In addition, it details the considerations for supporting each component of the Laserfiche system, from the Windows Server, through every Laserfiche application, to storage and backup.

Specifying and selecting hardware is more an art than a science. There's no single answer to any given question, and hardware needs are never the same from company to company. Though this paper contains a broad overview of the questions to ask and the decisions to make, there will be many variations in customer setups that will require you to ask questions and make decisions not covered here.

> **Note:** The estimates and advice in this paper are intended as guidelines only, not as rules. We cannot guarantee the hardware we suggest will always be sufficient for every organization's configuration.

Gathering System Requirements

The first step when planning the implementation of Laserfiche for an organization is to determine what the customer (or organization or department) expects from Laserfiche.

The most important question involves the user base. When considering your user base, remember that there is a difference between the total number of users and the number of users who might be logged into the system—and therefore using your hardware resources—at any single moment.

Because the number of active, concurrent users is incredibly important when considering appropriate hardware for an installation, we suggest an estimate of one-in-four—that is, for every four total users (licenses), one user is likely to be accessing the system at any given time. For a 200 user license enterprise implementation, you'll want to plan for your system to accommodate 50 simultaneous users at any given time.

> **Note:** Throughout this paper, we will use the term "users" to refer to simultaneous, concurrent user load on the system. This term used here should not be confused with the number of named user licenses or the previous Laserfiche licensing schema of "concurrent user licenses."

If your customer, organization, or department will deploy Laserfiche over the Web (e.g., Web Access or WebLink), take note of the number of users who will use Laserfiche this way, in order to calculate the amount of traffic that will be accessing your Web servers.

After getting a sense of the user base, the next step is to determine their expected scanning volume. How many documents will be going into the system on a daily and monthly basis? We recommend trying to predict or calculate this number for a three to four year timeframe (considering three to four years to be the typical lifespan of computer hardware). This figure should be in number of pages (storage implications and how to convert this number to GB are discussed below).

On a related note, you'll need to determine if image capture will primarily be done through an installed Laserfiche Scanning client on the same local area network (LAN) as the Laserfiche Server or if it will be done via Web Scanning. This information will help you calculate Web server loads, as well as any effect on network bandwidth.

Workflow and Quick Fields can be challenging applications to specify hardware for. However, for your initial requirements gathering, it's enough to know if the customer plans to use the products, how intense that use might be, and what, if any, Quick Fields add-ons they might employ.

System Storage Calculations

The table below provides guidelines for the amount of processing power and memory that might be required for an installation based on the number of concurrent users and the number of pages brought into the system each day.

- **The green area** represents user needs that can typically be supported by a single server running Laserfiche, Web Access, and Microsoft SQL. We suggest at least a quad-core processor with 16GB of RAM.

- **The pink area** represents needs that are best met by a dual-server configuration, with Laserfiche and Web Access on a single server, and Microsoft SQL on its own machine. For the Laserfiche/Web Access server, we recommend at least quad-core processor with 8GB of RAM, and for the SQL server, the same processor with at least 8GB of RAM.

    o With a system of this size, organizations may consider virtualizing the Laserfiche infrastructure. We recommend allocating four processor cores and 8GBs of RAM to a single virtual machine hosting both Laserfiche and Web Access servers.

    Note: For the best possible performance, keep Microsoft SQL on its own physical machine, even if other applications are virtualized.

- **The blue area** represents configurations approaching "enterprise" level, for which we recommend a three-server configuration: Laserfiche, Web Access, and Microsoft SQL each on their own machine with a minimum of a dual-core processor and 8GB of RAM.

> **Note:** With more than 100-150 concurrent users, a second Web Access server is recommended to support network load balancing.

- For virtualization, we recommend a similar configuration where each server (except for MSSQL) resides on a dedicated virtual machine. The virtual machines should be allocated four processor cores and 8GBs of RAM.

- **Optimum storage configurations** also depend on the size of the system. For smaller Laserfiche implementations, such as the green area below, we recommend a single RAID 5 array. For the other estimated loads (pink and blue in the chart above), the ideal configuration is three separate RAID arrays, one each for the Laserfiche volumes (RAID 5), Microsoft SQL Server data file (RAID 1 or RAID 1+0), and Microsoft SQL Server transaction logs (RAID 1 or RAID 1+0).

Active Concurrent Users

PAGES/DAY 50kb ea	10	25	50	100	200
200	15GB*	30GB	45GB	60GB	75GB
500	30GB	45GB	60GB	75GB	90GB
1,000	60GB	75GB	90GB	105GB	120GB
2,000	105GB	120GB	135GB	150GB	165GB
3,000	150GB	165GB	180GB	210GB	240GB
5,000	210GB	240GB	270GB	300GB	360GB
10,000	450GB	495GB	540GB	585GB	660GB

*Please note that all storage estimates listed are for three years, which is the estimated life of current off-the-shelf hardware.

When dealing with a truly enterprise customer and system, there are some best practices to follow. Putting all Laserfiche components on separate servers is well worth the extra cost to avoid potential resource-sharing problems. Also, don't forget additional support environments when you plan for Laserfiche in production. Development, testing, training, and production environments may all be required for seamless performance, depending on your customer's requirements.

As you create a hardware plan for your customer, plan your hardware system to include multiple cores and sockets. This will allow you to take advantage of Laserfiche's multi-threaded processing capabilities, making your customer's operations run more efficiently and quickly.

Evaluate Current State

Hardware and Network

After gaining a thorough understanding of what the customer expects from their Laserfiche implementation, you should survey the current system, not only to understand what resources the customer may already have in place, and what they may need to upgrade, but also to be sure you're aware of existing processes and behaviors you need to accommodate.

To start, look at your customer's existing network infrastructure, beginning locally with the LAN. Evaluate network speed. If your customer has remote sites, how are they connected? By a wide area network (WAN) or by virtual private network (VPN)? If remote sites connect via a WAN, be sure to evaluate that network speed, which will factor into a number of decisions. Do remote sites use a desktop client to connect across the WAN to the Laserfiche Server or do they use the thin Web Access client? If using Web Access, will there be enough users and network load to require more than one Web Access server? How is the installation distributed to remote users?

Note the specifics of the hardware in place, including the primary hardware vendor, as your customer may want to maintain existing contracts and/or relationships. Similarly, consider existing arrangements for storage, as enterprise customers usually have some sort of storage area network (SAN) in place. Again, access from remote business sites or partners will have an impact on retrieval times from storage. Another consideration is the customer's high-availability requirements: they may have service-level agreements for disaster recovery that must be met, governing failover locally or from site-to-site.

A final consideration, based on the current state of the hardware in your customer's system, is planning for growth. Laserfiche typically plans for three to four years of growth, which is considered acceptable life expectancy for off-the-shelf software.

Software

Laserfiche will not only need to work with existing hardware and networks, but also with existing customer software. Therefore, knowing what kind of software is already in use is also vital. What operating systems are supported? What database systems are in use already or are preferred? What backup software is being used with the customer's backup hardware configuration? You'll want to determine if your customer is planning to, or already does, virtualize their systems, and what virtualization software they use or prefer.

Planning

Hardware needs for each application component will be considered separately.

Windows Server

For the Windows Server 2008 Standard edition, the maximum server resources supported are four-processor sockets and 32GB of memory. For the Enterprise edition, the maximum server resources supported are eight-processor sockets and 2TB of RAM.

Laserfiche Server

Generally speaking, having more cores is better for performance than having faster processors. This is particularly true for the Laserfiche Server, which is a multi-threaded application and can make use of multiple processors or cores.

Concurrent users	Cores	RAM
200-400 users	8-16	8-32GB
400-1,000 users	16+	32GB+
>1,000 users	Contact reseller or Regional Manager	

SQL Server

When specifying hardware for your Microsoft SQL Server, we recommend allocating a separate physical machine. This protects other component resources and confine SQL to only the resources you designate. Internal Laserfiche tests have shown SQL performs better on a physical box than on a virtual machine.

Physical machine requirements will vary depending on Laserfiche load estimates, as well as the version of SQL Server. The SQL Standard version supports four sockets for processors and the maximum memory supported by the operating system. The Enterprise version, which you may want to use due to its support of failover clustering (a potentially important requirement for larger Laserfiche installations), supports more sockets and memory, up to the maximum number supported by the operating system.

Hardware Recommendations

The numbers required for the Laserfiche SQL database fall within the limits of the Standard and Enterprise SQL versions.

Concurrent User Load (user access, not user licenses)	Recommended Configuration
200-1,000 users	16-32 core, 32-64GB of RAM
1,000+ users*	32+ core, 64GB+ of RAM
*See your Laserfiche reseller or regional manager for more details or guidance.	

A best practice when planning and purchasing your system is to invest in multiple processors. In-house tests performed in two environments (one with a SQL Server with a single Intel Xeon x5365 processor, and another the same but with two processors) show similar search durations if there is only one concurrent search. But if more users log in to the system to search at the same time, the system with more processors is faster.

Another best practice is to buy fast processors. Searching, indexing, and full-table scans all benefit greatly from increased processor speed. In a test performed with an in-house system versus a production system configured by Laserfiche, we observed the difference in search result speed that a single gigahertz of processor speed can make. Both systems had identical specifications, except that the production site had a 2GHz processor, and our test system had a 3GHz processor. As the graph below indicates, the in-house test bed (red bars) returned the same search results in half the time of the production system (blue bars).
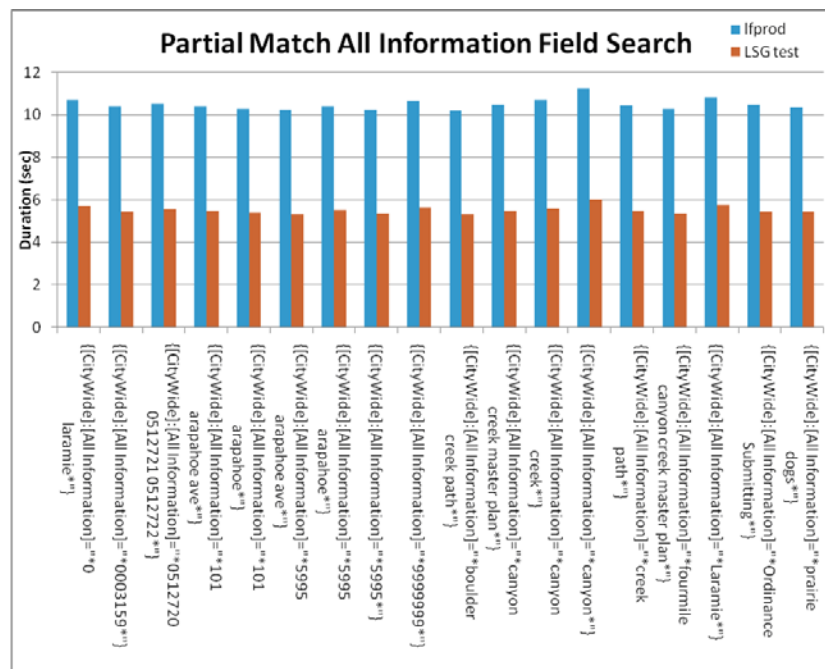
Another consideration with regard to your SQL server is how to handle multiple databases. In the Laserfiche Suite, Workflow, Audit Trail, and the Laserfiche Client/Server will each require a separate database, though they don't all have to be located on separate boxes. However, the Audit Trail database in particular will benefit from being separated onto its own SQL Server, due to its resource-intensive reporting function. Some organizations with Laserfiche even replicate all Laserfiche audit data into a separate reporting environment to minimize interference from other system demands and to maximize performance.

> **Tip:** Through customer site tests and our own in-house tests, we have determined that buying the second-fastest processor available may actually be the best practice. The second-fastest processors can cost as much as 20-30% less than the fastest processors on the market, and the speed differential is typically not substantial enough to warrant the premium.

Some customers may choose to utilize an existing SQL server installation for Laserfiche SQL needs. Up to a certain number of concurrent users and/or system load, this can be a cost-effective solution. We recommend a dedicated Laserfiche SQL installation for more than 50 concurrent users. Beyond that, we advise consistent monitoring of SQL usage and loads, to ensure enough resources are allocated for optimum Laserfiche performance.

Storage

The type of storage system connected to a Laserfiche implementation doesn't matter; whether it's a storage area network (SAN), direct attached storage (DAS), or network attached storage (NAS), Laserfiche can be configured to store volumes on any drive able to be referenced by a UNC storage path. Most enterprise systems, however, establish a SAN dedicated to storage needs.

The key metric for estimating storage capacity is how many images the system will house. As shown in the table below, Laserfiche recommends estimating 75GB per one million images, where each image is a single-page, 8.5x11, 300dpi, in black and white. The amount of storage space required will increase in ways that aren't easy to predict when working with large and/or color images. Storage required for SQL data and log files can be estimated at a percentage of total volume storage.

Recommended configurations for storage arrays are also detailed in the table below. For volume storage, we recommend at least a RAID 5 array, specifically for its redundant data storage (parity), which employs at least three disks to protect against total failure of any one. Note that for RAID 5 arrays larger than 2TB, we recommend RAID 6, with two parity disks, due to increased possibility of a failed disk. For SQL data and log files, we recommend either a RAID 1 array, for its mirroring function, or a RAID 1+0 array, which distributes data storage and mirrors data across four or more disks.

9

	Sizing calculations	Storage arrays
Laserfiche volume storage	75GB per 1 million images (8.5x11, 300dpi, black/white)	RAID 5 Serial Attached SCSI recommended Serial ATA OK
SQL data file	10-15% of total volume storage	RAID 1 or RAID 1+0 Serial Attached SCSI recommended with fast drives
SQL log file	5-10% of total volume storage	RAID 1 or RAID 1+0 Serial Attached SCSI recommended with fast drives
tempdb	5% of total volume storage (for larger Laserfiche installations, usually >200 users)	Shared with SQL log file

Note that the ratio between the SQL data file and the transaction log can vary greatly between "simple" and "full" recovery modes. In simple mode, the transaction log will stay quite small, and as per the table above, we recommend nearer to 5% of total volume storage for the SQL log file. In full mode, however, the transaction log can become very large, depending on how often you perform backups, and we recommend closer to 10% of estimated total volume storage.

Backup

Most organizations will have backup procedures in place, so planning for a Laserfiche system backup has more to do with determining the potential amount of data that will need to be backed up in addition to the current load. If a backup system isn't robust enough for the additional load or isn't completely set up, Laserfiche recommends a disk-to-disk-to-tape process: backing up to disk first—which can usually be the SAN, NAS, or DAS device—and then backing up the disk to a rotating set of tapes. If storing on a DAS, we recommend a large RAID 5 array with a number of SATA disks, or at least enough capacity for one month's worth of grandfather-father-son rotation. Larger systems should use LTO-standard backup tapes. Smaller systems with less data to back up can use smaller-capacity DAT drives.

Some enterprise deployments of Laserfiche use snapshot functionality available as standard hardware with disk backup to streamline the processes of backing up large data stores.

Other Component Servers

Web Server

More cores can mean better performance from your Laserfiche Web Server because of the way ASP.NET (the .NET host for Web-based .NET applications) handles input/output threads in IIS. RAM usually isn't as important or as relevant as the number of cores, with the exception of support for certain processes, such as PDF generation (for printing or exporting).

Laserfiche Web configuration offers unique flexibility in the Laserfiche system. With the Laserfiche Web server, as the load on the server increases, you can scale out and add more Web servers, even after initial implementation. To support Web-based load balancing, you must be able to establish server affinity to incoming client requests at the TCP/IP level. An example of this is Microsoft Network Load Balancing, which will dynamically allocate requests across the Web to multiple Web Access servers in a cluster.

Hardware Recommendations

Concurrent User Load (user access, not user licenses)	Recommended Configuration
<150 users	8 cores, 8GB of RAM
>150 users	More Web servers, network load balancing

Workflow Server

Laserfiche Workflow is a challenging product to specify hardware for, because the resource use of Workflow is determined by the number of workflows running, not the number of users. Moreover, workflow running speed is limited by the number of threads Windows Workflow Foundation can run.

Part of the load on the Laserfiche Workflow system is predictable; we can reliably estimate how many users are configuring workflows and starting rules in the Workflow Designer. From a planning perspective, Workflow most impacts the system when system actions are performed by workflow instances being triggered. It's hard to know how resource intensive an individual workflow will be, when creating it, much less when planning for a future system.

There are some guidelines and tips to follow for maintaining Workflow performance. Installing your Workflow Server in a virtual environment allows you to add (or remove) cores as necessary as you monitor production performance. In addition, using a machine with a faster processor with many cores allows the Workflow Server to take advantage of as many threads as the Windows Workflow Foundation can handle.

Hardware Recommendations

Concurrent User Load (user access, not user licenses)	Recommended Configuration
<200 users	4-8 cores, 4-8GB of RAM

For Laserfiche Workflow in particular, we recommend ensuring your test systems are similar to your production systems, so you have the best data possible for system load and resource allocation.

Quick Fields Agent Server

Though Quick Fields is a client-side application, it should be addressed in hardware plans because it will work the hardest of any of the Laserfiche applications, and it is very resource intensive. Faster processors will improve speed, particularly of popular, processor-intensive add-ons, such as OCR.

Hardware Recommendations

Quick Fields Product	Recommended Configuration
Quick Fields (uses up to four threads)	4 cores, 4GB of RAM
Quick Fields Agent (baseline)	8 cores, 8GB of RAM

Note: Quick Fields sessions can be run in parallel.

Since the number of processors or cores can be adjusted as needed, you might want to consider virtualizing the Quick Fields Agent environment. For storage sizing, start by allocating a minimum of 15GB in the virtual machine for the operating system and applications, plus 50KB (single-page, 8.5x11, 300dpi, in black and white) multiplied by the number of images to be scanned across all sessions.

Audit Trail

Audit Trail can be tricky to specify hardware for; its impact on the system depends on the events being logged. Audit Trail reporting can be an extremely resource-intensive process. In addition, Audit Trail event tracking can generate huge amounts of data, which can compete for storage space with other applications.

As a best practice, if you know that Audit Trail will be an important part of your system, you should replicate audit logs into their own reporting system in a completely separate environment. Many customers have found this to be the most efficient means of maintaining performance across their Laserfiche installation.

Conclusion

As you plan for a new Laserfiche installation, estimate hardware needs generously. Hardware needs will only increase as time goes on and as resources are added to the system. Once hardware has been purchased and put into production, you should continuously monitor the performance of products and operations across your system and add or reallocate resources as required. If you need help, ask for advice from Laserfiche and other customers.

Laserfiche®