Distributed by:

techinfocenter.com

Production Imaging

PRODUCTION IMAGING

Written and Produced by The Rheinner Group

Distributed by techinfocenter.com

TABLE OF CONTENTS

The Rheinner Group

The Rheinner Group is a leading research, consulting and education firm in the document imaging, management and workflow industry. Its Certified Document Imaging Architech (CDIA) Education Program, which covers many of the same issues addressed by the Rheinner Guides, is the most popular training program in the imaging industry. For more information on The Rheinner Group, CDIA course schedules, or to obtain help designing and implementing document imaging and workflow systems, please call 781-741-8100 or visit our web site, at www.rheinner.com.

The Rheinner Group is a proud sponsor of the Certified Document Imaging Architech™ (CDIA) certification program.

What Is Production Document Imaging?

All document imaging systems have common characteristics: First, they all involve the conversion of paper-based information into a digital format, called a document image. Second, they require the integration of subsystems to accomplish the capture, storage, retrieval, viewing and output of the resulting document images. Beyond that, document imaging can be deployed in a variety of environments, ranging from a setting in which users require access to a low-to-moderate volume of documents to an environment in which users require constant and uninterrupted access to a high volume of document images. The latter type of system is referred to as a production imaging system. At its most fundamental level, a production imaging system converts substantially higher volumes of paper-based information to digital format, and therefore requires more robust subsystems, peripherals and application software to perform the various functions associated with the capture, storage, retrieval, viewing and output of those document images.

Characteristics of Production Imaging Systems

Beyond this general definition, there are a variety of factors which indicate the need for a production imaging system:

- A high volume of document images: All other things being equal, production imaging systems tend to have high volume requirements. This means that there is a large volume of paper that comes into the department or organization each day, generally in excess of 5,000 pages per day, that daily retrieval requirements are high, and that thereare a large number of documents in the repository (generally approaching one million per year).
- Users spend a high percentage of their work time on the system. Production imaging system users tend to spend most of their time working with the system, and their work is highly dependent on the document images and the flow of work associated with those images.

- The system is transaction-oriented. Most production imaging systems automate well-defined repetitive and consistent business transactions, such as insurance claims processing, accounts payable or mortgage processing. Generally, the system user handles a large number of items, spending a small amount of time (typically less than 10 minutes) on each item. And the items tend to be similar.
- The system is mission critical. Production imaging systems typically replace key paper-based systems that are central to the organization. Processing this work is critical to the business because money, customer service and product delivery depend on these systems.
- The system may have a large number of users. Production systems generally have more users than do non-production systems. However, the number of users is the least important indicator of a production system. In a small company, the number of users of a production imaging system tends to be small, while in a large company the number of users tends to be large. Regardless of the number of users, if everyone in an organization uses the system, it is probably not a production system, while if most people in a formerly paper-based department use the system all day long, it probably is a production system.

Production systems can be enterprise-wide in that universal access to the document image repository is available, but an enterprise imaging system is not necessarily a production imaging system nor must a production imaging system be an enterprise system. For example, access by all company personnel to document images on an occasional and irregular basis does not constitute a production imaging system. And the majority of production imaging systems are departmentally-based designed to automate the functions of a single department. That's because while information can be ubiquitous in an organization, paper rarely is (with the minor exceptions of company-wide newsletters and similar things). The same smart organizations that build enterprise computer systems to automate both departmental and cross-departmental functions in a uniform and integrated way will strive to limit the distribution of large volumes of paper to a single department. Therefore, those progressive companies' needs for a production imaging system, while possibly involving very high volumes, are still likely to be limited to single departments.

Most companies' needs for a production imaging system are strictly driven by the volume of paper received and retrieved. For example, if you are a small business that only receives a couple dozen checks a day in the mail, the cost of handling them manually adds up to so little actual cost and effort, that it isn't worth even thinking about spending the money to automate the process. However, if you are a utility company receiving tens of thousands of checks a day, you have probably already spent money installing an automated remittance processing system, though perhaps not an imaging-based one. The larger the amount of work and the cost of processing the work, the more the opportunity to install a production imaging system to automate it.

Production vs. General Office Imaging Applications

Another type of document imaging system is sometimes referred to as "casual imaging", for lack of a better name. Also called "workgroup imaging", "ad hoc imaging", or "office imaging", this type of imaging system is designed to provide access to a wide range of document images for a broader range of business applications. In this guide, we'll refer to this type imaging application as "general office imaging". These systems support ad hoc information sharing to enable better decision making, and streamline general office functions, such as expense approval, budget approval, sales administration and customer service. These systems are typically characterized by low to moderate volumes of incoming documents and document image repositories, highly variable and unpredictable work processes, and low to moderate, but similarly unpredictable, document image retrieval requirements.

There is clearly a relationship between the degree to which an imaging application can be tuned and adjusted in order to achieve the highest possible throughput and efficiency, and the degree to which it can be put to use by normal users with minimal training, setup and customization. The higher the volumes and the more repetitive the operations, the higher the premium on the efficiency of the end result and the ability to achieve it. The lower the volumes, the more varied the operations, and the less time the users will spend with the system, the higher the premium on ease of use, ease of setup and quick customization.

Production imaging systems seek to achieve maximum efficiency and reliability, and as a result are frequently harder to set up than other kinds of systems. Production imaging systems place substantial demands on the computing platform, the imaging application and on each of the imaging subsystems and components.

Key Requirements for Production Imaging Systems

While all document imaging systems perform similar operations on images, production imaging systems have a set of requirements that go well beyond the capabilities of systems not designed with production use in mind. These requirements include:

- Reliability No one wants a system to break. The question is, how likely is a break down, and if it happens, what is the time and cost to repair, and how much value (lost customers, etc.) is lost during the down period? In production imaging systems this is a key question, which is not always given the consideration it deserves.

In anything other than a production system, you calculate the throughput you need from the scanner, and buy accordingly. In a production system, you should calculate your throughput if one or more of the scanners has broken. In other words, you should spend more time planning for the case you hope never happens (knowing that it sometimes will) than you do for the "normal" case.

- Disaster Recovery Backup is another issue that is often overlooked in the average imaging installation. In a production imaging installation, backup of everything is a given. What if there's a flood? What if there's a fire that takes out the main production facility for weeks? What if power is cut for a couple of hours to the location where two hundred people work? It doesn't take long, first to calculate the financial consequences of events like these, second to realize that what used to sound like overkill sounds cheap compared to the consequences of a disaster. The best planned systems are subject to loss because of people; think of the billion dollars worth of damage (including the irreparable loss of millions of documents) due to the collapse of the tunnel underneath the Chicago River. This disaster was actually predicted well in advance. The government director in charge tried to get it fixed, but ended up mired in a bureaucratic tangle of procurement regulations to assure that no one made too much money preventing this billion dollar disaster!

- Throughput Lots of systems have components which can, in bursts, reach or exceed throughput rates of large-scale production systems. But throughput is not about components. It is about whole systems, from paper handling through drivers, from networks through file systems. Having components, each of which have the apparent capacity to meet production imaging throughputs, is just the starting point. Making them work together, each at optimal performance, is where the real work begins.

Unfortunately, it is not just a question of choosing the right components. It is a question of testing everything together under realistic conditions. Most people find the results to be surprising, typically disappointing. “We deliver the bits to the interface, but the bus can’t take them fast enough.” “We had to write to that driver to conform to the standards, but it wasn’t designed for overlapped I/O, so the actual throughput is only a third of the rated speed.” “That particular release of the operating system doesn’t give control back soon enough for us to catch the rotation on our really fast drive, with the result that it’s actually slower than drives rated half its speed”. The caveat: Don’t rely on demonstrations, rely on live tests with realistic volumes, actual proposed equipment and software when designing and implementing a production imaging system. Vendors don’t offer to submit to such tests lightly, and indeed sometimes actually must be paid to do it. But what is the cost of the time lost trying to make a system work which ultimately fails? Live tests prior to installation are the prudent equivalent of disaster recovery—spending money to assure that a possible disaster does not take place.

- Mainframe Considerations In spite of all the talk about client/server systems, there is no more mature and reliable system than the mainframe. The vast majority of large-scale businesses continue to run their key operations on mainframes. Mainframes set the standard for reliability, security, integrity, and smooth operations. While the majority of production imaging systems do not run on mainframes, mainframes continue to set the standard for what “production” means in computer environments; even when there is not a mainframe in sight, it is wise to consider mainframe standards when designing a truly mission critical imaging system.

Performance Issues in Imaging Platforms

Production imaging is all about reducing human labor, achieving high quality and consistent results, and reducing cycle times to a bare minimum. These objectives are achieved when an imaging system delivers high performance on a consistent and reliable basis.

Throughput vs. Response Time

It is easy to confuse throughput with response time. In fact, the two factors are often at odds with each other. General office imaging systems, particularly those shown on trade show floors, typically have sensational response times. Response time is measured from the time a user makes a request until the time the request is satisfied. The typical example is picking a document from a list and measuring the time it takes to display the document on the screen. This example makes sense for many document imaging installations. However, for most production imaging systems, it is irrelevant.

In production imaging, there is typically a large number of documents which have to be processed in the course of the day. Some portion of the new documents need to be matched with other documents from the archives in order to be processed. Does it make sense to scan a new document, route it to a human being, have the human decide what archived document is required, request it, and wait for it to show up on the screen? In any reasonably serious production imaging installation, the archive document is likely not to be available for immediate recall to the screen. It is likely to be stored on a medium, disk or tape, which is not mounted in a drive. A jukebox or operator will be required to fetch and mount the tape or platter for reading. Delays of at least a few seconds are likely, even assuming that the drive and jukebox in question were idle at the time of the request. In short, having operators request documents from a large archive on an ad hoc basis is not likely to lead to satisfactory results in a production imaging system.

In production imaging systems, the key idea is "staging," accomplished by a carefully planned series of batched operations, many of which are performed by programs running in the background, on a server machine. With staging, the document images that will be required for work in the future (typically the next day) are moved from optical or other storage devices onto local magnetic disk (typically during the night). The result of staging documents: excellent response times. Every single operation which is staged requires thinking, planning, setup, and therefore, cost. Staging is not a significant component of general office imaging systems since the occasional five-second delay to a user is not too problematic. But multiply a few seconds times hundreds of operators times hundreds of operations per day, and the delay becomes significant and counter-productive. In this scenario, batching and staging operations help the system deliver optimal performance.

Perhaps the best example of the response time vs. throughput trade off comes in fetching documents from optical platters onto a magnetic disk cache. If requests for 100 documents stored on 20 platters come randomly and are processed immediately in the order received, on average about 95 platter exchanges will have to be performed by the jukebox. If a platter is loaded in the drive and a new request is processed, the odds are only five percent (one in twenty) that the new document will be on the platter that is already in the drive. So it is swapped out and the required drive loaded in. But what if all 100 requests are given (in batch!) to a smarter fetching program, which is optimized with minimizing total throughput in mind? The program will sort all the requests into the order of the platters on which they reside. The program will then load each platter in turn, fetch all documents from the 100 requests that reside on that platter, replace it and go onto the next. A maximum of 20 platter swaps will be required to read in all 100 documents, compared to 95 swaps for the method which optimizes response time, nearly a 5:1 advantage. As the number of documents increases, the differential grows. Consider 1000 documents. The response time method will require on average about 950 platter swaps, while the throughput method will still only need 20 swaps, nearly a 50:1 advantage.

Application Integration

Another hallmark of production imaging is application integration. An excellent example of this is to contrast a prime goal of general office imaging—on-the-spot OCR integration—with the legacy system integration so often found in production imaging installations.

In general office imaging, the user's fax software running on his PC receives a fax. The user calls up the fax software and views the fax. The user decides that this fax is interesting, and copies the fax to his imaging software. After the import, he requests OCR. The systems does its work. Naturally, errors are detected. The user views and corrects each error in turn. The application then deposits the results into Microsoft Word. For the individual user, this entire process is infinitely preferred to the alternative, which is printing out the fax and key entering it into the word processing application. However, it is actually still quite a time consuming process.

In a production imaging installation, the fax is received on an unattended fax server, a machine with multiple fax boards. The new fax is routed to an application which classifies it using forms recognition, performs image processing, removes the form and recognizes the text. Each individual field of the form that cannot be recognized is routed to an operator, who is presented with just the error fields, at heads-down key entry rates of speed. After correction, the data is assembled and passed to a program for further processing. The total amount of human intervention is seconds; but the amount of time required to initially setup the program is possibly days, compared to almost none for the previous scenario.

Not only are production imaging applications high throughput, they are typically focused on extracting actionable data from documents which arrive in large numbers. Some of the processing will be done by the new image-centric application, but much of the data processing burden falls on other applications. While the most common of these applications are legacy applications that have been up and running within the organization for many years, frequently document management and general groupware applications such as Lotus Notes are also involved. Interacting with these external applications in an efficient and reliable manner is key to the performance and acceptability of the system.

Subsystem Stress Points

Production imaging systems bring out the worst in the components used to build these systems, because the loads placed on the system push each individual component to its maximum.

The most vulnerable spots are the places where subsystems interface to each other. The storage system, for example, may perform well in unit tests, but when it is integrated with the rest of the system, things may not go as well.

This well-known effect is simply due to the isolation of the designers of each component from each other. This is actually an argument against the current trend towards component imaging. While this trend is inevitable in many ways, in high-end production systems, it requires additional attention from systems integrators to assure that all components in fact perform together in spite of the fact that the engineers who built them may never have been exposed to the peculiarities of other components. This fact also reinforces the user's need to rely on live tests using actual volumes of actual documents to evaluate a system's performance.

Workflow Requirements

Production imaging systems are very likely to involve significant integration with workflow to manage the document capture and data extraction process automatically, and to automatically route and prioritize the work associated with the document images. Workflow can be used to help achieve optimal system performance, for example pre-fetching other documents associated with a case so that a user has all the documents required to adjudicate the newly-arrived document.This is classic, document-centric workflow automation, which can only be built after a thorough understanding of the document flows is achieved.

Work management is another goal of workflow processing software. One of the most important functions of workflow software is controlling and measuring the flow of work through an organization. Even if that work amounts to little more than distributed, image-based data entry, only by tracking all the work performed at each workstation is it possible to identify and correct problems and to spot opportunities for improvement.

On the other hand, few factors have been more instrumental in delaying the installation of production imaging systems than the specter of re-engineering and workflow analysis. Workflow can provide an excuse for endless rounds of discussions, idealistic-sounding theories, and graphs which are tantalizing and elusive at the same time. If there are benefits to a potential production imaging system, there are costs to delaying its installation. Workflow is a good thing. Workflow analysis which delays the installation of a perfectly adequate production imaging system which will bring benefits, is not a good thing.

Implementing Production Imaging Systems

As mentioned earlier, a production imaging system is harder to configure than other imaging systems, and places substantial demands on the computing platform, the imaging application and on each of the imaging subsystems and components.

The Input Subsystem

Successful design of the input subsystem is a make-or-break issue for production imaging systems. Failures of input mean that none of the workers on the system get a stream of work to do. Seemingly minor inefficiencies of input result in bad or unreadable documents, or in unacceptably high costs.

While it is tempting to leap right into equipment and technologies, a good production imaging input design starts with an exhaustive and painstaking analysis of the documents. The documents have to be broken into types, and everything about each type needs to be identified: how to recognize the type, its page count, the kind and placement of information on each page, each piece of information which needs to be extracted. Then the arrival rates and locations must be determined over time, including an analysis of weekly, monthly and seasonal variations by type.

Samples of stacks of paper input should be examined by experienced experts who can devise a plan for document preparation. A number of important questions have to be decided, not only the handling of the documents in preparation for scanning, but also regarding their disposition after they are captured by the system. It is wise to set up a strict scheme of batching, with numbered and dated batch headers and trailers. If the paper is to be saved in off-site storage (an easy and inexpensive way to eliminate concerns about digital storage), the database has to maintain the correspondences between documents, batches, and boxes.

In production imaging, the process of feeding the documents to the scanner is normally semi-automated, which means that the scanner takes documents that are fed to it individually by a human operator. Only when the documents are extremely uniform and in very good condition is it possible to consider a completely automated document feeder.

There are many details about scanners which need to be considered, too many to be included in this guidebook. You should assume that, in practice, there will be a primary scanning station that is always capturing documents, but a flatbed scanner with good controls should be included to handle the exceptions. You should assume that the scanner will break, and plan for how to get it repaired and what to do in the interim. You should consider detailed features of the proposed scanner in the light of your plans. For example, a scanner with a straight-through paper path is the most conservative choice for paper handling. However, if you plan to reassemble the documents after scanning, a straight-through paper path may result in the sheets being dropped on top of each other at the back end of the scanner, which means that they will end up in the reverse order.

The technologies of image processing have advanced dramatically in the least several years, fueled both by increased processor power and improved software algorithms for processing. These algorithms can be applied inside the scanner itself, inside a control board, in an attached processor, or in the host machine which controls the scanner. Certain image processing operations have to take place in a particular order. For example, the ability to pick out very faint writing against a light background or writing on a colored or shaded background requires access to all the gray scale values of the original scanned image, which are normally discarded by the time the image gets into the controlling server. Therefore, this process (typically called adaptive thresholding) is nearly always a feature of the scanner itself. Other operations are better performed in the server. Other valuable image processing operations include edge detection and cropping; skew correction, which makes images take less space, more readable, and increases the rate of automated recognition; automated flipping, which corrects the operator's feeding documents in the wrong orientation; and forms recognition.

After documents have been captured, data has to be extracted from them. Two main steps comprise data entry: indexing to capture the data needed for later retrieval of the document, and data entry to capture the data needed by the application programs. These functions can be done in two separate steps or at the same time. The most efficient method is to extract all data that will be necessary by anyone or any program in a single step.

There are a variety of methods of extracting data from images. In production imaging system, great emphasis is placed on automating the extraction and verification process, because much of the cost and quality of the system comes from this step. In rough order of precedence, the following techniques are used for data extraction:

- Bar code recognition is highly effective because the bar code is either immediately recognized or not recognized, like the beep on the supermarket checkout counter. The number of false positives (the machine thought it read the bar code properly, but in fact it did not) is minimized.
- Forms recognition applies algorithms to extract the document type from the appearance of the form.
- Mark sense recognition sees whether a check box is filled in or not.
- Image character recognition applies complex algorithms to attempt to read characters on the document. The more constrained the physical area where characters may appear, and the more cross checks that can be applied to check the accuracy of the recognition, the better the results.
- Human key entry can be used for original key entry, for double-key verification, or to correct or check the results of previous methods.

Human key entry is nearly always part of the data extraction process. The question is, how well designed is the entry— the objective is presenting the key entry person with only the work that can't be done by other methods in a way that maximizes his or her efficiency.

The Display Subsystem

The requirements for image display are stringent in production imaging systems because the people who use the system are likely to spend all day looking at the display. Displays that look fine in a demonstration room can lead to eyestrain when used all day long under less than optimal lighting conditions. More than for any other imaging system component, consideration should be given to the human factors and the actual conditions in which the display will be used.

When users have to read whole documents, which is typically the case for documents that are not forms, having a large display that can show the documents in at least their normal physical size is indispensable. If the users spend most of their time with the documents themselves and little time with other applications, a gray scale display is generally preferable, because the practical resolution of the image is greater, and it is more readable. If the readers spend time with mixed applications along with images, most people prefer to accept somewhat lower image resolution in order to get color with their non-image applications.

During production-quality image key entry, a smaller color display is normally preferred. This is because the software will generally only show the part of the image which needs to be keyed, typically a small part of the whole, blown up to greater than life size. The power of color can be used to guide the user in his work.

Communications

While many general office imaging applications can be added to existing LANs without difficulty, a production imaging installation can stress an organization's communication environment. The principle of communications with imaging is straightforward. Each network has a certain capacity. Some of that capacity is made unavailable because of overhead, and some is used by existing applications and users. This leaves a certain amount of available capacity. This is a number that can be measured by an experienced network administrator. The question is: will the demands of the imaging system exceed the available capacity of the network?

In order to determine the answer, you have to first lay out the topology of the proposed image network. There are sources of images (scanning workstations and storage servers) and destinations of images (storage servers and user workstations). Each image has a measurable size, typically in the range of 30KB to 120KB per image. Sources send images to destinations at a certain rate, which can be predicted within reasonable limits prior to installation. Multiplying everything out gives the required network capacity. This calculation should include allowance for overhead, peak loads and momentary surges; accounting for additional network transfers required by the document capture process; and backup and disaster recovery requirements.

Industrial-strength scanners can scan in excess of 100 pages per minute, which is nearly 2 images per second. If each image is 100KB, that is already a load of 200KB/sec, compared to a theoretical Ethernet capacity of about 1MB/sec (in bytes, not to be confused with the usual Ethernet specification of 10Mb/sec, megabits per second). Consider the following scenario: If each image goes over the network three times for storage (once for a temporary cache server and once to each of two storage servers, each backing up the other), the requirements reach 500KB/sec. If each image is then fetched four more times, twice for indexing (once for automated recognition and once for a human being) and twice for workflow requirements, network requirements climb to 900KB. This is at least twice the comfortable load on an Ethernet that was empty prior to use by the imaging system, and makes no allowance for file server and database server traffic.

There are solutions to this kind of capacity problem. One frequently used strategy is to isolate all the activity related to scanning, image processing and data extraction onto its own subnetwork. In this case, the storage server has dual connections, one to the document capture subnetwork and one to the workflow subnetwork. The workflow network itself can also be divided into regions of isolated activities. However, this has to be done carefully to exploit the patterns in the workflows, but avoid limiting future flexibility in rearranging the work.

Storage

Storage is one of the most hotly discussed topics in imaging generally. In production imaging, where the stakes are often higher and the investments required often greater, the heat can become intense. But there are a couple of core ideas which apply to the analysis of storage requirements independent of the storage media and devices that are used.

The most important concept is that of the storage hierarchy. Take a given amount of storage, for example, 1GB, which represents about 20,000 images. Those images can be stored on a variety of media and in a variety of devices. Choosing which device and media becomes a tradeoff between availability and cost. The more quickly you want to be able to access any of these images, the more expensive the cost of that storage. Conversely, you can reduce the cost of storage by accepting longer delays in storing and retrieving the images.

The top of the storage hierarchy is normally magnetic disk, whether it is arranged in single spindles or in banks under the control of a RAID subsystem. RAID may be required to increase the throughput of the central cache of active images, for example, at some stage of the document capture process. RAID may also be desirable to increase the availability of the images, protecting against single disk failures. After the magnetic disk, the choices get more complicated. When moving images to the next stage of the hierarchy, the process is usually optimized if treated as a batch process; for some media (for example, CD-R and tape), batch write is a practical necessity. In any case, the writing is normally done on servers in the background, while the document management system directs all requests for the documents being stored to the copies stored on magnetic disk.

Historically, the medium of choice for secondary storage of images has been write-once optical disk (WORM). This choice is mandated by regulatory requirement today in some cases, and in other cases remains an excellent choice, providing the best access times for unanticipated, random reads. Production imaging installations, in which response time and throughput is at a premium, tend to prefer this medium. However, depending on the extent to which fetch requirements can be anticipated in advance and other needs analysis, other media such as rewritable optical disks, tapes of various kinds, or writable CD-ROMs may be appropriate. Certainly CD-R has been growing in acceptance in the last couple of years. Once created, depending on retrieval requirements, the media may be placed in direct drives (best response time, most costly), jukeboxes (next best response time, medium cost) or passive shelving (human-assisted media loading, worst response time, very low storage cost). Passive shelving is nearly always used, for example, for off-site secure tertiary media storage.

Migrating documents among the various media can also get complicated. It is essential in production imaging that some kind of system-managed storage be applied to achieve location transparency so that the user can request a document without regard to its storage location. In addition, system-managed storage should automatically migrate documents from expensive, fast-access media to less-expensive, slower-access media when the frequency of retrieval of a document drops off.

Architectural Issues

Early production imaging systems met their performance and functionality goals by means of specialized hardware and software. But the closed and proprietary nature of these systems prohibited the integration of new more capable components, as they evolved. With today's production imaging systems, even those on the largest scale, it is possible to integrate carefully selected hardware and software components into an existing computer environment and still achieve the required production characteristics. While integration of specialized components will always be part of the production imaging environment, users can demand that even these components conform to industry standards and interface conventions.

The simplest level of integration is with applications. The document imaging system typically needs to interact with other applications. It may need to link documents to accounts that are maintained by another application. It may need to look up things in an existing application. Most commonly, it may provide data that must be put into data-entry screens of existing applications. While it is possible to place the burden of doing this on operators using a split-screen method where the operator copies the information from one side to the other, it is more common in high-volume situations to go to the extra effort and create one unified approach. Using development tools like Visual Basic, it's possible to create a single user interface that integrates the two application.

Sometimes it is possible to bypass the application and go directly to the underlying databases, particularly if these are built using a relational DBMS. This makes it possible to add fields needed to integrate the image documents to existing tables without breaking existing applications.

Integrating with existing platforms is nearly always a requirement. Depending on the circumstances, it may make sense to run the production imaging on newly purchased servers, particularly the image storage servers. However, every imaging system needs a database for document cataloging, and using a database on an existing platform for this purpose may make sense.

Upon analysis, many production imaging systems use documents which are entirely internal in nature. These documents may represent requests for services made by one department to another, for example. In some cases, the online imaging system may simply eliminate the need for internal documents. Frequently, however, the need is not eliminated, because not everyone who needs to send or receive such a form has access to the imaging system. This is a case where integration with external messaging or groupware systems is essential, so long as the volume is not too large for the external system to handle. For example, the legal department may need to know when a certain document has arrived, and can be notified of the arrival by a message. It is best, of course, if the image itself can be attached to the message.

Similarly, workflow events may need to cross departmental boundaries. This is most easily accomplished if everyone in the organization is part of a central workflow system. But this may not make sense, since production workflow systems tend to have relatively large per-seat license fees. Therefore, programming a transition between a certain workflow workstep and an external system may be a good idea.

Finally, integration with other application environments can be used for exception processing, to move “problem work items” out of a production imaging cycle, where they can slow things down, over to personnel who can resolve these issues. For example, in a production accounts payable imaging application, an invoice may need to go back to the user department because of a discrepancy in the amount billed. That user may not have access to the production imaging database; but might be a Lotus Notes user or have a CC:mail box, for example. A gateway from the production system to these applications enables the production user to send the document image to that user for problem resolution.

Image Management Application Development

Even well-planned production imaging systems do not pop out of the box and work. They need some level of programming to achieve the throughput and efficiency goals. While integrating the system with external applications is an obvious need for programming, some combination of careful parameter setting, control table setting, database setup and programming is usually needed before the system can become operational.

For example, the document types have to be set up, queues established, index fields defined and equipped with edit checks and formatting, storage migration rules have to be established, case management has to be established and controlled, pre-fetching has to be integrated, and various document rendezvous need to be arranged for. Some of these items involve interaction between the imaging system and the workflow system. For example, the workflow system establishes the need to wait for the arrival of a document, but it is the imaging system that sees the new document arrive.

Implementing Production Imaging: Rules of Thumb

Making a production imaging system operate to specification on the shop floor is a challenging exercise. It can only be achieved if each individual component is capable and has been sized to be able to handle the projected load.When components have not been designed from the start to work with each other, they may work more poorly together than either of them does in isolation. This surprising result is frequently seen in real life situations. The reason it arises is simple: each device has a model of how the device on the other side of the interface is supposed to act. If two devices have conflicting expectations, no good will result. Even worse is when two important components don't have the same interface; then, special means need to be taken to make them work together, and can be a source of tremendous inefficiency.

The message is that the way to make systems perform well is first to make the parts perform well in isolation, and then to make each pair of interactions perform well, keeping on until the entire system performs as required. This critical requirement is the main reason why practical experience with production imaging is an important vendor selection criterion.

Vendor selection for production imaging involves all the factors applied to selecting vendors in general. While no one wants to spend too much money, factors such as reputation, financial stability, experience, products which meet the requirements, etc. have to be taken into account.

However, it is important to remember the differences between production imaging and other forms of imaging. The fact that wonderful and practical applications of document imaging can be effective in low-end environments due to the power of today's computers and networks can easily seduce you into thinking that a few two thousand dollar Pentiums, a scanner and jukebox, and a little software will make a great system. Actually, it can make a great system—just not one that meets any reasonable definition of production imaging on anything but a modest scale.

So the importance of selecting vendors who truly understand what production imaging is all about, and have demonstrated that understanding by building systems that work, cannot be stressed enough. This applies to vendors at each level—hardware components, software modules, system planning, integration and installation.

System Design Guidelines

While general office imaging systems concentrate on what happens at the client workstation, production imaging systems are server-heavy. The design allows for lots of things to happen in the background, for example redundant servers or mirrored systems at geographically separated locations. This background focus should also include efforts to minimize human time with the system by anticipating user needs. For example, work can be moved to workstations in anticipation of its being needed. As a result, networking analysis concentrates on throughput, not instantaneous response. Finally, given the mission-critical nature and high cost associated with most production imaging installations, system design should concentrate on the long view. A screen that looks great for a ten minute demonstration could lead to a lawsuit because of employee problems due to eye strain.

Perhaps the biggest difference between system design for general office systems and production systems is need for full scale testing prior to production installation and use. Many aspects of a plausible design simply can't be predicted, and only by trying the system out under realistic loads can the real issues be surfaced and corrected.

Device drivers—Programs that tell the computer how to communicate

Author Biography

The Rheinner Group is a leading research, consulting and education firm in the document imaging, management and workflow industry. Its Certified Document Imaging Architech (CDIA) Education Program, which covers many of the same issues addressed by the Rheinner Guides, is the most popular training program in the imaging industry. For more information on The Rheinner Group, CDIA course schedules, or to obtain help designing and implementing document imaging and workflow systems, please call 781-741-8100 or visit our web site, at www.rheinner.com.

Copyright Information

techinfocenter.com

techinfocenter.com

Rheinner Guides are educational pieces written by techinfocenter.com analysts. The booklets are sponsored by leading vendor organizations eager to educate and inform users about the capabilities of new technologies and applications.

Written By: **The Rheinner Group**