

Image-2

PROJECT

Image-2: Document Processing System

IMAGE-2

DOCUMENT PROCESSING

Project Image-2

© Project srl
Via Gramsci • 50019 Sesto F.no (Firenze)
Tel. 055.0515465 • Fax 055.0516392
www.project-sistemi.com
info@project-sistemi.com

Image-2: document processing

1. IDP functional requirements

The functional requirements for Image-Document Processing (IDP) systems usually start as simple Archiving Systems, but need to be modularly upgraded to a true Document Management System able to manage compound documents in their native format.

This because:

- ◆ OA Systems are unable to really improve the productivity of the document management process.
- ◆ Usual Filing Systems, such as microfilm or microfiches, are not enough flexible, mainly because their difficulties in document distribution and retrieval.

Productivity can be really improved only integrating Image-Document Processing functionality with the OA ones, and the main IDP functionality are so as follow.

1.1. Documents Acquisition.

Documents acquisition (which means scanning, indexing and commit) has to be characterized by the following functionality.

- a. Quick, accurate and automatics document scanning in one of the main standard image format (TIFF, PCX, etc)
- b. In order to maximize the efficiency and the efficacy of the process, all the document acquisition operations are to be opportunely organized depending on the operator demand.
 - ◆ Scanning document by document or in batch, for large document volumes with automatics or manual page-document separation.
 - ◆ Doing all the steps (scan, indexing and commit) on the same stations or on different stations.
 - ◆ Interrupting and Resuming any steps in any time.
 - ◆ Splitting all the document acquisition operations in attended steps (scanning and indexing) and in unattended ones (commit, file transfer).
- d. Automatic image conversion, indexing and commit of documents electronically produced via the most important Windows based word-processor and spreadsheet (Word, Excel, Lotus, etc), via VT100/3270/5250 ascii file transfer with a mainframe or via printer spooling using COLD techniques.

1.2. Document-Image DB

The Document-Image DB has to be able:

- a. To archive documents in their original format quickly and efficaciously, with the ability:
 - To automatically assign a Document-Identification in order to univocally identify each document in the DB for improving security issues.
 - To assign each document type to an end-user defined Document-Template, which states elaboration rules (security for instance) and retrieval indexes for all the documents which will be assigned to that Document-Template.
 - b. To use DB media standard, based on, up to date technologies, modularly expandable and able to maximize document integrity and security.
 - c. To use different DB in order to improve system security:
-

Index-DB, based on magnetic disks, in which are stored all the document indexes and the Document-Templates, managed by one of the standard RDBMS.
Archive-DB, based on optical disks, Warm or multifunctional, in which are stored document images.

1.3. Documents Retrieval.

Documents retrieval has to be characterized by the following functionality:

- a. All the LAN workstations can contemporary and securely access the Documents-DB using various retrieval techniques:
 - In real time for analytical query, with one or more access keys, or in batch unattended for large statistical reports.
 - In browsing on sets of documents with one or more attributes defined via standard operators (binary, SQL, etc).
- b. Various level of security and privacy.
- c. Ability to contemporary access both local Documents-DB and others corporate DB, via VT100/3270/5250 or other terminal emulator.

1.4. Documents manipulation and distribution.

The operator, directly at the workstations, has to manipulate and to distribute document's images and data.

- Display, rotation, zooming etc, without any updating of the original document image.
 - Print on printer server or on local printer.
 - Document indexes update.
 - Document annotations and/or amendments.
 - Interaction with corporate Mainframe (3270, 5250, VT100, etc) via inquiry on the corporate DB or downloading-uploading files with the local DB.
 - Document faxing and mailing, both locally or outside the company.
-

2. IMAGE-2.

IMAGE-2 is based on a client-server architecture, both at Hw and Sw level, where it integrates the most up-to-date and standard Hw-Sw components oriented to Image-Document Processing.

IMAGE-2 is based on the main Windows interprogram communication mechanisms (DDE and OLE) and programming environment (Visual Basic, Access, SQL interface, etc).

IMAGE-2 it's LAN and DB independent and it's able to evolve from a stand-alone solution to a complete LAN solution.

2.1. IMAGE-2 basic concepts.

Some of the most important criticisms of Image-Document Processing solutions are related to flexibility and operational issues, and IMAGE-2 is able to resolve most of these issues because its systematic application of the 3 main following concepts.

a. **Doc_Id.**

It is an identifier which univocally identifies each IMAGE-2 document and which is automatically assigned by the system, in order to improve both document retrieval and document security.

So, for instance, the system, automatically and for each Doc_Id, is able to writes operator's name and date of the last document's indexes update.

b. **Session.**

Session is an unit of job composed of more than one steps.

The job of course ends only after the last step, but all the steps can be done one immediately after the other or, *viceversa*, the user can stop a step and he can after resume it or the next step at any time, without any job degradation.

In IMAGE-2 the documents acquisition (scanning, indexing and commit) can be done step by step or not, so IMAGE-2 gives to the operator the possibility to organize its work depending on its specific needs.

Ability to cut all the job in attended steps, such as scanning and indexing, and in unattended steps, mainly commit and printing, which can be done automatically and during the night.

Ability to stop and, on user demand, to resume scanning, indexing, assembling and printing of documents with a lot of pages.

Ability to add or to update other pages to documents already scanned and indexed.

c. **Document_Class.**

Each document has to be managed with specific rules which are different depending on document functions inside the organization; so normally in an OA environment it's usual to define Document-Templates, which exactly mach this Document-Class IMAGE-2 basic concept

Document-Class is used in IMAGE-2 in order to predefine application rules and indexes for all the documents which will be assigned to the same Document-Class.

Because the systematic application of this concept, IMAGE-2 is able to improve the system's security and privacy level:

- by reserving specific archiving volumes for specific Document-Class
- by adding to the normal file-server access control mechanisms, someone specific for each Document-Class.

2.2. IMAGE-2 SW Architecture.

Because IMAGE-2 is Windows based, it's perfectly integrated with all the main Windows based software products, and it's open to evolve for integrating new Windows based software and hardware technology.

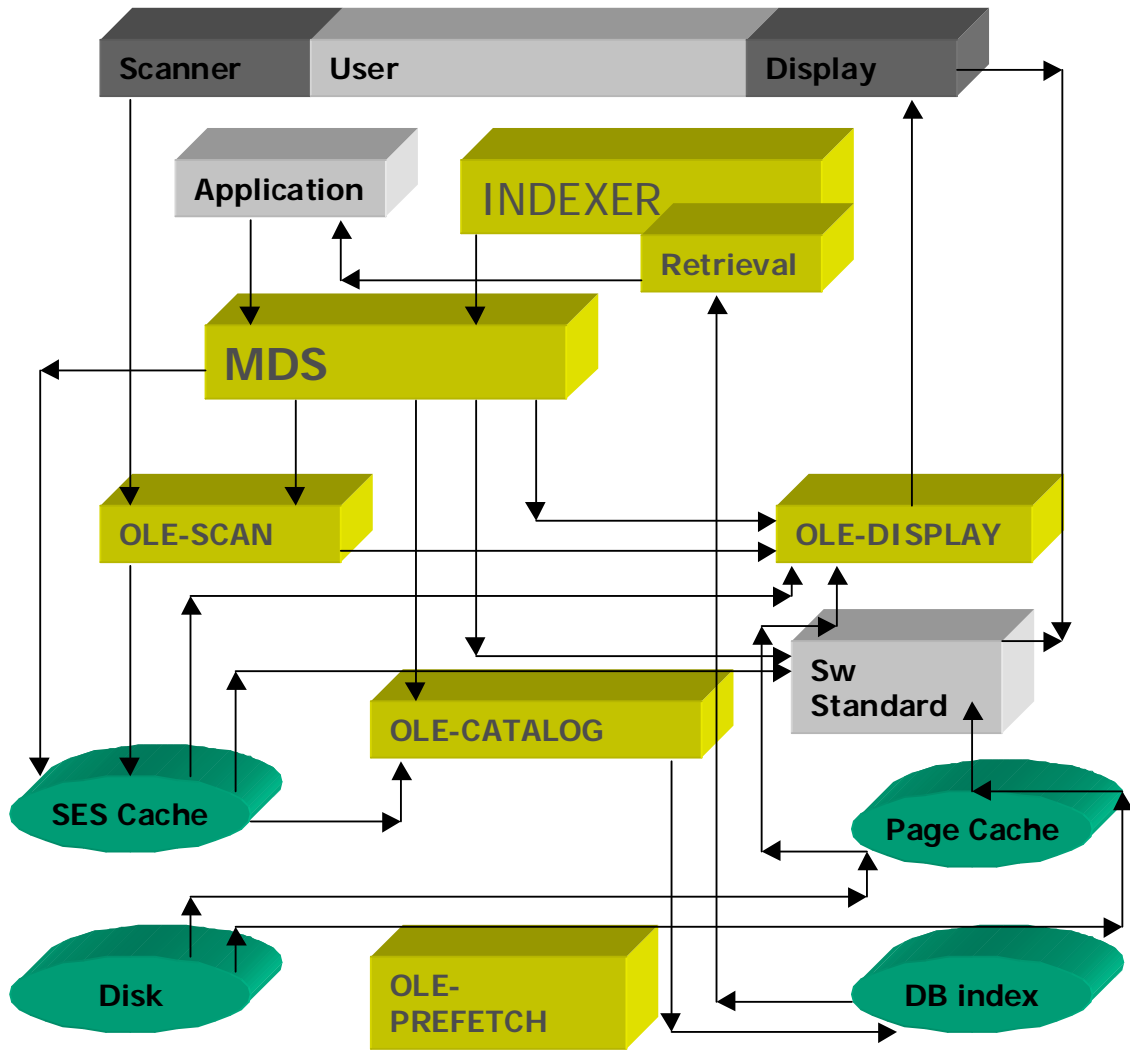


IMAGE-2 software is modular and it is composed of OLE servers and of so

me application modules which operate as clients.

a. MDS-Server.

It is composed of an Index-Server, which manages and stores document indexes (OleCatalog), and an Archive-Server, which manages and stores document images.

b. OleDispl-Server.

Image&Document Processing Server implements all the main document management functions required by the applications.

Display, rotation, zooming, conversion of document images acquired via scan/fax and compressed-decompressed in the main image format (TIF, PCX, etc), with compression-decompression usually done on board, for performance issues.

Management of all the sessions of print, fax and scan, both interactive and batch, using all the most important caching techniques.

c. Indexer.

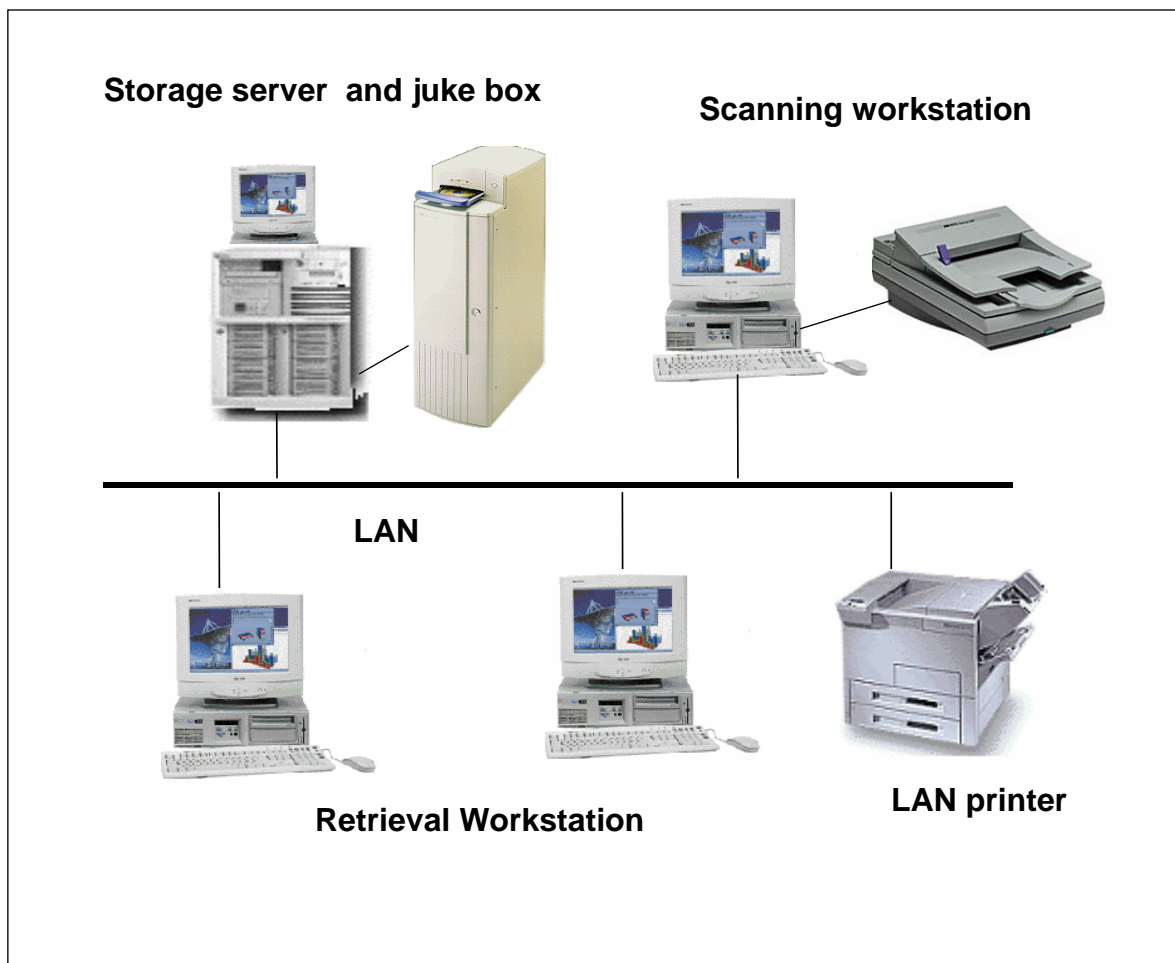
Each application module operates as Client and it interacts with the IMAGE-2 Ole Servers from one site and with the operator from the other site, in order to do all the above functions for retrieving, scanning, indexing, commit and document manipulation and distribution.

Because IMAGE-2 is based on standard Hw-Sw platform (Windows i.e.), because its client-server architecture and because its above pictured software modularity, it is able to insure:

Total independence from Lan cabling (Ethernet or Token-Ring) and file-Server (Netware, Lan Manager, WindowsForWorkGroup, etc).
Perfect integration with other Windows software products, such as Mail, Word, Excel, Lotus and with all Windows based user applications.
Upgrading from a stand-alone to a Lan IMAGE-2 configuration and/or with new up to date hw-sw technologies.
Easy migration on future sw platforms, such as Windows NT for instance.

2.3. IMAGE-2 LAN configuration.

The IMAGE-2 Lan general system configuration is defined in the picture below.



MAGE-2 functionality are modularly software implemented, so the difference between an IMAGE-2 stand alone solution and a IMAGE-2 Lan solution is only because in the first one all the functionality are integrated in the same station.

a. IMAGE-2 MDS Server.

MDS *via* OleCatalog relates indexes and Doc_Id of each document to its images, via two different server.

Used for storing Documents indexes and Document-Classes (see below), it can use any DB such as ACCESS, Oracle, Informix etc.

Used to store document images, it is usually implemented with storage server directly connected to the station in IMAGE-2 stand alone system, or to the Lan File-Server in IMAGE-2 Lan.

b. IMAGE-2 Client Stations.

The application functionality of each IMAGE-2 Lan station are implemented by integrating different software modules:

Ole-Display

It is an OleServer, it manages all the functions, for improving performances, and all the image manipulation functions, driving specialized compression-decompression engine (both hardware and only software).

It is an only software component installed in all the stations for Display-Retrieval, Scanning, Printing, Faxing, Mailing etc.

Function module.

It is an application specific sw module which interacts as OLE client with IDP-engine module and which is optimized:

- in **Image-2-D&R station**, for display-retrieval and acting as front-end to the other IMAGE-2 server stations for asking document print, mail and fax.

Optionally, and if there are not print performance issues, IMAGE-2 gives the possibility to print document directly into a local and parallel laser-printer.

- in **Image-2-Scan station**, for scanning.

c. Specialized Image-2-Server.

IMAGE-2 of course uses also other server types, such as:

Image-Mail/Fax Server.

IMAGE-2 client integrates front-end functions for Microsoft-Mail software, hence:

- In order to implement Fax-Server functionality, it is sufficient to use MsMail and install one of the country and Microsoft-Mail certified FaxGateway.

- In order to implement a Mail-Gateway for external E-Mail systems, it is sufficient to install one of the MsMail and country certified MailGateway.

2.4. System RAS.

As far is concerned this issue, we have to underline the following main IMAGE-2 characteristics.

a) System Reliability and Availability

The system RAS is based not only on the File-Server back-up functionality, but also on customized recovery-restart procedures, which will be possible to better design only after a more accurate application analysis.

Anyway, because the session approach of Image-2, the system implements easy and robust recovery-restart mechanisms based on session step by step status journaling hence, via this journaling, IMAGE-2 is able to automatically align itself to the last well captured document's page.

b) System Security

This issue is based both on the file-server functionality (for which please see the file-server specifications) and on the following special IMAGE-2 standard mechanisms, which are able to improve the system security and which can be managed only by IMAGE-2 System Administrator.

◆ User-identification

During IMAGE-2 login each user has to identify itself to the system typing its own User_Id, which is checked by IMAGE-2 via an internal table with the User_Id of all the authorized IMAGE-2 users.

Hence only authorized users can go ahead in the IMAGE-2 login.

◆ Authorized user-operations.

As far is concerned the operations authorized to each user, during the login IMAGE-2 recognizes 3 main kinds of users:

- Supervisor.
He can do all the possible IMAGE-2 operations.
- Scanning user.
He can do only documents scan, indexing and commit
- Retrieval user.
He can do only documents retrieval.

Hence, during the login, IMAGE-2 is able to activate, on the Windows menu of each user, the functionality which implement only its authorized operations.

◆ Document access control.

Each IMAGE-2 Document-Class is related with different passwords, which are to be declared by the user, also if supervisor, each time he opens a Document-Class.

- Only users who know Password can open the related Document-Class.

◆ More granularity in controlling document's access.

All the above features are standard in IMAGE-2 and usually they are enough for the security requirements of a lot of Document-Management applications.

Anyway some applications need much more application-dependent granularity in controlling document's access, so in this cases it's necessary to build-up customized controls which, of course, can be done only after a more specific application analysis.

2.5. Other important IMAGE-2 specifications.

Other important issues are related to programmability, document capture and document annotations.

a) Customization.

IMAGE-2 is easily customizable using its API to interact with its OLE-Server. Hence its services can be requested, via API, by any Windows program which can operate so as a OLE client and it is very easy to integrate it in any Windows sw product.

b) Electronic Document.

Usually it's useful also to automatically capture, to convert in TIF and to store in the Document-management DB also all the electronic documents, which are internally produced via word-processor (WinWord, etc) and spreadsheet (Excel, Lotus, etc).

IMAGE-2 provides standard WinWord and Excel macro in order to automatically capture, convert in TIF and index all this kinds of document, and, on user request, it is possible to develop the same functions for any other Windows based software packages.

c) Document distribution.

IMAGE-2 already integrates full MsMail functionality using standard Microsoft MAPI, hence it is still ready to implement document mail/fax distribution functionality, which can be expanded outside the company simply adding on the LAN some Microsoft certified gateway (fax, X400, EDI, etc).

d) Document annotations

IMAGE-2 stores as ASCII file, on its DB also document's note and amendments related to each document's page, hence it is able to display, contemporary and into separate windows, each document's page with all its own related notes.

e) Mainframe interaction

This will be realized via standard package VT100, 3270, 5250 etc terminal emulator such as Rumba or ExtraForWindows and, if necessary, will be also possible to integrate it inside the IMAGE-2 document management functionality.

2.6. IMAGE-2 configurability guideline.

a. LAN

All the most important Lan Operating Systems (Novell, NT, UNIX, AIX IBM) with all the types of Lan wiring and adapter.

b. Fax

All the most important fax adapter CAS compliant, standard CCITT group III and IV, with modem Class 1 and 2.

c. Compression/Decompression Engine

KOFAX, with more than one types of hardware engine and also with only software engine, in order to optimize the cost/performance ratio of all the types of stations.

d. Scanner, Printer, Optical disk, Monitor.

All the main models are supported.
