

**I**l presente capitolo affronta una delle tematiche più delicate legate ad «Immagina» ed all'archiviazione documentale in generale: il corretto dimensionamento delle risorse di sistema sia in termini di periferiche, sia di configurazione hardware del server e dei client della rete. Sicuramente le problematiche che emergono nel presente capitolo non vengono approfondite come dovrebbero (non sarebbe sufficiente un intero volume dedicato all'argomento), ma certamente l'obiettivo di far soffermare il lettore su alcune considerazioni viene raggiunto pienamente. Ci scusiamo in anticipo per la sensazione di approccio ingenuo che il lettore più preparato potrebbe trarre da queste pagine, in realtà l'utilità di questi brevi accenni è assodata dal numero di richieste in tal senso che è giunta al nostro supporto tecnico in passato. In considerazione del tenore di tali domande, dunque, si è scelta la strada dell'approccio informale e discorsivo. Sugeriamo a chi avesse già dimestichezza con tali argomentazioni di scorrere velocemente i vari paragrafi.

### 21.1 Considerazioni generali

Chi legge ha quasi sicuramente già acquistato «Immagina» (ottima scelta!), di conseguenza ha già considerato come dimensionare, più o meno adeguatamente, le risorse da dedicare all'attività di gestione documentale. Ciò nonostante riteniamo utile approfondire l'argomento, in quanto l'occupazione di memoria dei documenti è legata a parametri talmente variabili da creare situazioni inaspettate per l'amministratore che non abbia ben compreso quanto detti parametri incidano a seconda delle scelte adottate. Non sono rare

purtroppo, nel campo dell'archiviazione documentale, situazioni in cui lo spazio calcolato per la memorizzazione dei documenti risulti enormemente inferiore alle reali necessità.

Nella sezione relativa alla scelta corretta delle periferiche di memorizzazione vengono forniti alcuni elementi che aiutino l'amministratore nella valutazione delle diverse possibili periferiche da utilizzare in funzione del tipo di documenti da memorizzare, della risoluzione e del formato. Raccomandiamo un'attenta lettura in particolare a coloro che prevedono di avere la necessità di memorizzare immagini a colori, in quanto l'occupazione di queste ultime può riservare sorprese (spiacevoli).

## **21.2 Periferiche di Input e Output**

In questo paragrafo ricordiamo semplicemente i principi generali che guidano le scelte delle periferiche di I/O. È necessario tenere sempre presente, nella scelta dei dispositivi periferici più adeguati, le prestazioni che si vogliono ottenere e la mole di lavoro che ogni singola periferica sarà chiamata a svolgere sia in media, sia in quei "picchi" di lavoro che, diverse volte nell'arco dell'anno, capita di raggiungere.

La periferica ha un carico di lavoro ideale, in funzione del quale è stata progettata e costruita, che il costruttore sovente indica esplicitamente nel manuale tecnico, o nel materiale illustrativo, esprimendola in pagine trattate al mese. In verità questo accade più di frequente per le periferiche di output (stampanti) che non per le periferiche di input (scanner) dove il parametro indicato è generalmente la sola velocità di acquisizione immagini espressa nell'unità di pagine al minuto, anche se in questi ultimi anni si sta diffondendo tra i produttori di scanner, non solo altamente professionali, l'abitudine di indicare la produttività ideale (espressa in pagine al giorno). È tuttavia necessario che si tenga presente che, come in ogni altra "catena informatica", l'anello più debole è rappresentato da quei dispositivi che, avendo prestazioni limitate dalle caratteristiche meccaniche, rischiano di diventare un "collo di bottiglia" per le prestazioni globali.

Banalizzando possiamo sintetizzare dicendo che se il nostro carico di lavoro sarà di diverse centinaia di migliaia di documenti all'anno, non potremo certo ipotizzare di acquisire tutte le immagini con uno scanner da poche pagine al minuto.

Uno degli aspetti su cui è necessario porre particolare attenzione è che, qualora si scelga come periferica di acquisizione dati uno scanner di fascia economica, c'è ancora, a distanza di oltre dieci anni dalla nascita della tecnologia a 32 bit, il rischio che il driver sia a 16 bit o che, comunque, contenga elementi sviluppati a 16 bit. Questo, potrebbe comportare problemi nella gestione di applicazioni concorrenti durante le fasi di scanning, con una conseguente riduzione delle prestazioni operative della postazione di lavoro.

### **21.3 Documenti ed occupazione di memoria**

Abbiamo già avuto occasione di sottolineare in altri punti del presente manuale che l'occupazione di memoria di un documento è un dato estremamente variabile. I parametri che influenzano tale variabilità sono: il tipo di file dell'immagine, il formato, la risoluzione, la modalità (cioè a colori, a toni di grigio o in bianco e nero) e soprattutto l'efficacia dell'algoritmo di compressione. Quest'ultima voce, in particolare, rende ogni indicazione possibile sull'occupazione di memoria dei vari tipi di documento estremamente sommaria. L'algoritmo di compressione è molto più efficace quando agisca su pagine "semplici", quali ad esempio le fatture, piuttosto che su pagine più "complesse", quali ad esempio, copertine di riviste. Questo fatto risulta intuitivo se si pensa a quanto spazio inutilizzato, e quindi facilmente comprimibile, è normalmente presente su una fattura o su un documento di trasporto e a quanto poco ce ne sia su una copertina di una rivista.

Premettiamo che le dissertazioni del presente paragrafo interesseranno ad una minoranza degli amministratori di «Immagina», in quanto, anche se le possibilità operative del prodotto sono molto varie, la maggior parte lo utilizzerà per archiviare documenti standard (in formato A4, in bianco e nero a 200 o 300 dpi).

Ciò premesso è necessario che si accenni all'occupazione di memoria di un documento in vari casi, in modo tale da permettere all'amministratore di quantificare le proprie esigenze o di comprendere come usi diversi da quello abituale potrebbero influenzare la saturazione delle risorse a sua disposizione.

Nello specchio riassuntivo che segue possiamo identificare l'ordine di grandezza della suddetta variabilità:

<b>Tipo di documento</b>	<b>Formato bmp</b>	<b>Formato jpg</b>	<b>Formato tif</b>
<b>DDT A4 b/n a 200 dpi</b>	<b>485 Kb</b>	ND	<b>35 Kb (G4)</b>
<b>Copertina di rivista A4 b/n a 200 dpi</b>	<b>485 Kb</b>	ND	<b>45 Kb (G4)</b>
<b>Copertina di rivista A4 a colori a 200 dpi</b>	<b>11.455 Kb</b>	<b>1.840 Kb (FQ10)</b>	<b>7.500 Kb (LZW)</b>
<b>Foto 13x18 cm a colori a 400 dpi</b>	<b>10.887 Kb</b>	<b>1.830 Kb (FQ10)</b>	<b>7.500 Kb (LZW)</b>

Lo specchio riassuntivo illustra l'occupazione di memoria di diversi tipi di documento: un documento di trasporto tipo, un foglio A4 (la copertina di una rivista) ed una fotografia a colori in formato 13 x 18 cm.

Nella prima colonna sono riportate le suddette tipologie di esempio con la relativa indicazione della risoluzione con cui sono state acquisite e della modalità a colori o in bianco e nero.

Questi casi sono stati distinti per tre diversi tipi di file di immagine: **bmp** (formato standard non compresso), **jpg** (con fattore di qualità 10) e **tif** (con compressione di tipo G4 per i

documenti in bianco e nero e compressione LZW per le immagini a colori). Ovviamente i tre scelti sono solamente alcuni tra i numerosi tipi di file disponibili, ma sono sufficienti a dare un'idea dell'occupazione di massima. Resta ovviamente inteso che «Immagina» non pone alcuna limitazione sui tipi di file memorizzabili.

La terza colonna non riporta alcun valore per i documenti in bianco e nero, in quanto **jpg** è un formato applicabile solamente ad immagini a colori.

Un ultimo dato interessante, non inserito in tabella, riguarda l'occupazione di memoria di un'immagine a colori in formato A4 con risoluzione di 600 dpi in formato **bmp**, che risulta di 103.095 Kb; come dire che in un disco da 1 Gb ci si potrebbero memorizzare al massimo 10 immagini di questo tipo.

Dallo specchietto si evince chiaramente che il formato **bmp** è improponibile per un uso standard, a causa dell'eccessiva quantità di spazio richiesta.

Il formato **jpg** rappresenta un compromesso tra qualità dell'immagine ed occupazione di memoria, in quanto la qualità ha un decadimento che è legato al fattore selezionato. Non è un caso che, grazie al notevole risparmio di spazio che offre, il formato jpg sia il più diffuso in ambito Internet.

Il formato **tif**, che per inciso è il più utilizzato per i documenti in bianco e nero in ambito di applicazioni di archiviazione documentale, offre una compressione molto efficace nel caso di documenti non a colori, ma sensibilmente meno efficace nel caso di documenti a colori.

Un aspetto che non riguarda propriamente l'occupazione di memoria, ma che è opportuno segnalare, è il rallentamento delle prestazioni generali al crescere della dimensione fisica dei file di immagini che vengono trattati.

Questo significa che acquisire, trasferire ed elaborare immagini a colori ad elevata risoluzione richiede notevoli potenze di calcolo; vale a dire processori veloci e notevoli quantità di memoria RAM.

## 21.4 Periferiche di memorizzazione

Il presente paragrafo non può e non vuole essere una guida alla scelta delle periferiche di memorizzazione, in quanto non è questa la sede per un confronto di prestazioni tra i diversi dispositivi disponibili sul mercato o per una analisi tecnica delle caratteristiche hardware delle diverse soluzioni.

In questa sezione si evidenziano semplicemente i criteri generali che guidano l'amministratore nella scelta di una tipologia di supporto di memorizzazione piuttosto che un'altra. Il ruolo strategico di tale scelta è evidente, dato che sulla tipologia di supporto scelto verranno memorizzati fisicamente tutti i documenti. Questo si traduce in ottime prestazioni globali se la scelta è stata effettuata a ragion veduta, prestazioni deludenti se nella scelta si è optato per una soluzione non ottimale.

L'approccio più generale prevede che nella valutazione venga dato il giusto peso ai seguenti elementi:

**Capacità di memoria:** è questo il fattore primario da prendere in considerazione. Con l'ausilio dei parametri indicativi del precedente paragrafo l'amministratore deve quantificare, di massima, le sue esigenze medie annuali di spazio; in funzione di questo valore dimensionare il supporto di memorizzazione. In sintesi l'amministratore, basandosi su valutazioni legate all'attività di archiviazione cartacea, deve individuare quanti documenti di ogni tipo avrà presumibilmente la necessità di memorizzare nell'arco dell'anno. Una volta ottenuto questo valore dovrà scartare le soluzioni che non soddisfino le esigenze di spazio individuate. Per esemplificare quanto detto supponiamo che un'azienda abbia la necessità di memorizzare solamente documenti contabili scannerizzati con una risoluzione di 200 dpi in bianco e nero con una mole prevista di circa 100.000 fogli A4 all'anno. Per quanto visto prima lo spazio richiesto sarà di circa 4 Gb all'anno. Con un ormai vetusto disco rigido da 40 Gb l'amministratore non dovrebbe avere necessità di dischi aggiuntivi per almeno 5 anni

(considerando un pessimistico aumento “fisiologico dell’attività” di circa il 20% ogni anno, senza prevedere estensioni ad altri settori aziendali).

Nel caso in cui l’amministratore avesse necessità di scannerizzare ogni anno circa 5000 fotografie a colori con una risoluzione di 600 dpi si troverebbe ad avere necessità di diverse decine di Gb ogni anno. In tale condizione la scelta sarebbe inevitabilmente orientata a periferiche quali array di dischi o NAS (fino a pochi anni fa sarebbe stato quasi obbligatorio rivolgersi a juke-box di dischi ottici o magneto/ottici).

**Prestazioni:** Con il termine generico prestazioni si è voluto raccogliere sotto un’unica voce diversi aspetti tra loro collegati.

Il primo aspetto riguarda la sicurezza dei dati. A questo proposito i dischi standard sono la soluzione che meno tutela l’amministratore da perdite di dati, ma che sicuramente ha il costo inferiore. I produttori di CD ROM garantiscono l’affidabilità del supporto per un numero variabile di anni, il che porta i CD ROM ed i juke-box ad essere una soluzione sicuramente più indicata rispetto ai dischi standard, ma per contro più onerosa sotto l’aspetto sia economico, sia operativo. Attenzione a non considerare i DVD alla stessa stregua dei CD, in quanto la loro masterizzazione con strumenti non professionali rende l’affidabilità nel tempo molto inferiore a quella dei CD. Gli array di dischi SCSI, o le unità a dischi direttamente collegabili alla rete, spesso sono dotati di sofisticati algoritmi che tutelano l’amministratore da perdite di dati (e.g. RAID) e permettono la sostituzione di eventuali dischi rotti senza che si debba spegnere la periferica; negli ultimi anni il costo di queste tipologie di dispositivi è sceso notevolmente.

Il secondo aspetto riguarda la praticità. Nel caso in cui non fossero disponibili in linea i documenti, ma ci fosse la necessità di inserire nell’unità un differente supporto, si possono creare dei tempi morti che, in taluni casi, potrebbero rappresentare un limite eccessivo. Su un singolo CD ROM, ad esempio, possono essere al massimo memorizzati 600 Mb di dati, il che è sufficiente a mantenere circa 15.000 fogli nel formato A4 b/n a 200 dpi. È probabile

che, se non si dispone di un juke-box, nel caso di una ricerca piuttosto vasta, si ottenga come risultato un elenco di documenti residenti su CD differenti. Nel caso in cui si volessero visualizzare tutti sarà necessario fisicamente alternare nel lettore di CD ROM i vari CD secondo le indicazioni fornite dal programma. In questo senso si può identificare la praticità di un supporto con la possibilità di non rallentare l'attività degli utenti.

Il terzo aspetto riguarda i costi. Tutti gli amministratori di «Immagina», si presume, sarebbero orientati ad adottare la soluzione tecnicamente migliore, qualora potessero, ma non tutti ne hanno i mezzi economici. Un juke-box ad elevate prestazioni può raggiungere costi di parecchie migliaia di euro. Questo lo rende una scelta giustificabile solo a fronte di necessità che non tutti hanno. Per contro, il costo sempre più basso (a fronte di capacità sempre più elevate) dei dischi magnetici rende queste unità, opportunamente backupate, una soluzione pressoché obbligata.

**Sviluppi futuri:** è necessario prevedere quale evoluzione avrà l'attività aziendale. L'esperienza ci insegna che più l'archiviazione elettronica diventa una risorsa aziendale, più aumenta la mole di documenti che vengono trattati ogni anno. Questo significa solamente che in una realtà aziendale è piuttosto raro che una soluzione di archiviazione documentale venga adottata direttamente per automatizzare tutte le attività aziendali. Di norma si procede per gradi. Ad esempio si può adottare in un primo tempo la soluzione per automatizzare il processo di archiviazione dei documenti di trasporto e delle fatture; in un secondo momento la si può estendere alla corrispondenza; in una terza fase si può completare estendendola a tutti i flussi cartacei aziendali. Nel calcolo delle esigenze è necessario ipotizzare il tipo di evoluzione che il processo di automazione subirà ed in che tempi ciò dovrà accadere. In tal modo il dimensionamento sarà adeguato, negli anni, alle crescenti necessità aziendali,

In sintesi possiamo confermare ciò che vale per ogni scelta in campo informatico, e non solo, cioè che la bontà della soluzione è legata a considerazioni soggettive. Si tratta di trovare, in ogni singolo caso, il giusto equilibrio tra costi e prestazioni. Nel presente capitolo

abbiamo accennato ad alcuni aspetti che l'amministratore deve conoscere al momento della scelta dei dispositivi di memorizzazione da acquistare. Non è compito del presente manuale fornire le basi di conoscenza per poter effettuare tali scelte.