

# CAP1

# ARCHIVI

**Ing. Daniele Corti**



copyright

all rights reserved

Copyright © Ing. Daniele Corti 2013

[www.ingdanielecorti.it](http://www.ingdanielecorti.it)

Tutti i diritti sono riservati a norma di legge e a norma delle convenzioni internazionali.

Ver.1.0

## **PREREQUISITI**

- ✓ Conoscenze sulla struttura fisica di un computer.
- ✓ Conoscenze sul concetto di file e memorizzazione permanente dei dati.

## **OBIETTIVI**

- ✓ Riconoscere le differenti tipologie di archivio.
- ✓ Riconoscere i concetti chiave di un archivio informatico.

## **ARGOMENTI**

- ✓ Gestione degli archivi.
- ✓ Archivio di dati.
- ✓ Dall'archivio cartaceo all'archivio informatico.
- ✓ File.
- ✓ Dato e informazione.
- ✓ Codifica e decodifica.
- ✓ Record – registrazione.
- ✓ Campo.
- ✓ Gestione degli archivi.
- ✓ I/O standard.
- ✓ Software di gestione dei file (cenni).
- ✓ Classificazione delle memorie (cenni).
- ✓ Sistema informativo.
- ✓ Sistema informatico.
- ✓ Dall'archivio informatico tradizionale alla base di dati.

# CAP 1 – ARCHIVI

## ARCHIVI CARTACEI → ARCHIVI DIGITALI → DATABASE

### GESTIONE DEGLI ARCHIVI

La funzione principale dei programmi gestionali è quella di memorizzare una grande quantità di dati su un supporto fisico, per poi recuperarli ed elaborarli in un secondo tempo.

### ARCHIVIO DI DATI

**Un archivio è un insieme organizzato di dati che vengono salvati su un supporto.**



L'archivio è un concetto astratto che pur se adattato a oggetti fisici anche molto diversi fra loro (scaffale con raccoglitori, struttura metallica con cassette) ha come scopo comune quello di mantenere raccolte le informazioni relative ad un determinato contesto.

L'uso degli archivi deriva dalla necessità di conservare i dati in modo permanente al fine di poterli utilizzare in futuro.

### DALL'ARCHIVIO CARTACEO ALL'ARCHIVIO INFORMATICO

Con l'avvento delle nuove tecnologie, a partire dagli anni 50-60, gli archivi cartacei (fisici) furono – piano piano – sostituiti da quelli informatici (digitali) ottenendo i seguenti vantaggi:

- Minore ingombro.
- Ricerche di dati più veloci.
- Durata più elevata.

- Fruibile da più persone contemporaneamente.
- Fruibile anche in remoto.
- Maggiore sicurezza.
- Utilizzo di archivi multimediali per memorizzare non solo testo ma anche disegni, foto, immagini, audio e video.



La registrazione dei dati su un supporto di memoria di massa deve essere fatta in modo da organizzare i dati nel migliore modo possibile, ottimizzando l'uso dello spazio occupato sul supporto e minimizzando i tempi di accesso e ricerca dati.

Anticipiamo che in un archivio le informazioni vengono memorizzate secondo insiemi di **record (schede)**. Un record è un insieme di informazioni fra loro correlate; per esempio per l'archivio Alunni possiamo avere diverse schede del tipo Mario Rossi 5A, Luigi Verdi 3B, ecc.

Le singole informazioni che compongono un record si chiamano **campi**; per esempio nel caso dell'archivio Alunni abbiamo i campi Nome, Cognome e Classe.

I dati memorizzati negli archivi possono essere gestiti attraverso diverse tecnologie:

- **Organizzazione sequenziale:** i record vengono salvati uno di seguito all'altro, in modo sequenziale; la ricerca di un record verrà effettuata scorrendo tutti i record del file a partire dal primo. Questa organizzazione è la più semplice esistente e deriva dalla classica sistemazione dei pacchi di schede o del accesso al nastro magnetico: per poter accedere a un record devo aver visitato tutti i record precedenti. Questa tecnica è sicuramente poco efficiente in termini di ricerca del dato soprattutto se l'archivio ha dimensioni elevate. Inoltre se si vuole aggiungere un nuovo record (non necessariamente di lunghezza fissa) occorre inserirlo in coda agli altri già presenti.
- **Organizzazione ad accesso diretto (casuale):** tutti i record dell'archivio hanno la stessa lunghezza, e in questo modo è possibile accedere direttamente ad un record sulla base della conoscenza della sua posizione all'interno dell'archivio. Ogni record, quindi, viene identificato dalla posizione che esso occupa nell'archivio.

- **Organizzazione a indice:** ogni record è identificato da una chiave. La ricerca di un record nell'archivio è, quindi, paragonabile alla ricerca di una pagina di un libro mediante il classico indice analitico. La chiave può essere una particolare caratteristica (uno o più campi) che identificano univocamente il record nell'archivio. Per esempio, nel caso dell'archivio Alunni la chiave può essere il codice matricola. La ricerca di un record (pagina) nell'archivio (libro) non verrà fatta scorrendo direttamente i record dell'archivio ma andando a leggere la tabella degli indici (indice analitico).

Gli aspetti principali della gestione automatizzata degli archivi sono rappresentati da:

- la tipologia dei **supporti** utilizzati per registrare le informazioni;
- le attrezzature **hardware** dedicate alla gestione delle unità di memorizzazione;
- gli strumenti **software** per la costruzione di programmi applicativi e dell'interfaccia per l'utente;
- la definizione dell'**organizzazione** degli archivi per rendere efficiente l'accesso ai dati e veloci le operazioni di ritrovamento.

Vediamo le proprietà di un archivio:

- i dati contenuti sono fra loro legati secondo una certa logica (inerenti ad un certo contesto);
- i dati sono rappresentati con un formato che ne rende possibile l'interpretazione;
- i dati sono registrati mediante un supporto su cui è possibile scrivere e rileggere anche a distanza nel tempo;
- i dati sono organizzati in modo da permettere una facile consultazione.

Esempio: l'elenco telefonico è un archivio di dati in cui le informazioni riguardano gli abbonati al telefono di una provincia; per ogni abbonato sono riportati nell'ordine generalità, indirizzo, numero di telefono; tutte queste informazioni sono stampate su fogli di carta.

Dal punto di vista informatico, un archivio può essere pensato come un file, e quindi possiamo definire il file nel seguente modo:

## **FILE**

**Il file è l'unità logica di memorizzazione dei dati all'interno di un supporto informatico.**

Il supporto è il mezzo fisico in cui sono fisicamente conservati i dati dell'archivio.

Fisicamente sul mezzo fisico verranno salvate delle sequenze di bit che rappresentano i dati contenuti nell'archivio.

Definendo l'archivio come un contenitore di dati, è naturale pensare che, come l'archivio, anche il dato è un concetto astratto, la cui definizione è:

## DATO E INFORMAZIONE

**Il dato è una qualsiasi informazione che si vuole registrare o memorizzare all'interno di un calcolatore.**

Gli aspetti caratteristici di un dato sono:

- Il **nome** che lo identifica (per esempio, data\_nascita).
- Il **valore** che può assumere (per esempio, "7 novembre 1968").
- Il **formato** ovvero la forma con cui il dato viene conservato in un archivio e ne rende possibile l'interpretazione (per esempio 07/11/1968).
- La **correlazione logica** del dato con gli altri dati (per esempio nome, cognome e data di nascita sono dati inerenti ad uno stesso argomento).
- Il **supporto** sul quale i dati vengono salvati e successivamente letti e riscritti anche a distanza di tempo.
- La **modalità** con la quale i dati sono organizzati in modo da permettere una facile consultazione.

Occorre però fare una distinzione fra il concetto di dato e di informazione:

- Il dato descrivere aspetti elementari di entità e fenomeni.
- L'informazione è un insieme di dati elaborati e presentati sulla base dell'esigenza di utilizzo pratico da parte delle persone interessate.

Quindi i dati rappresentano quegli elementi che vogliamo conservare in memoria o negli archivi, in vista di un successivo trattamento per produrre le informazioni, che forniscono una maggiore conoscenza della realtà sulla quale si intendono attivare operazioni di controllo, modifica o direzione.

Il trattamento dei dati per ottenere le informazioni viene indicato con il termine **elaborazione**.

## CODIFICA E DECODIFICA

Processi che consentono l'interazione fra macchina e uomo. La fase di codifica consente la traduzione del dato dal linguaggio naturale (dell'uomo) nel linguaggio del calcolatore. La fase di decodifica, quindi, è l'operazione di traduzione inversa dal mondo digitale al mondo umano.

## RECORD - REGISTRAZIONE

In un archivio, in genere, le informazioni sono raggruppate secondo un'unità logica (per esempio nel caso della gestione dei voti i dati relativi ad ogni alunno). L'insieme delle informazioni logicamente organizzate e riferite a un unico soggetto viene chiamato record.

**Il record è l'unità macroscopica delle informazioni.** Un singolo record può essere suddiviso logicamente in più parti. Per esempio, sempre nel caso della gestione dei voti, l'insieme delle informazioni logicamente raggruppate fra loro, per costituire un record, sono:

- Nome
- Cognome
- Classe
- Indirizzo
- Telefono

Dal punto di vista grafico, un record può essere pensato come la riga di una tabella:

Nome	Cognome	Classe	Indirizzo	Telefono
<b>Paolo</b>	<b>Rossi</b>	<b>4AT</b>	<b>Viale Roma, 30</b>	<b>02456789</b>

## CAMPO

Le parti logiche in cui viene suddiviso un singolo record sono dette campi. In genere, il nome del campo deriva dal tipo di informazione che è associata al campo stesso.

## GESTIONE DEGLI ARCHIVI

Vediamo quali sono le operazioni standard per la gestione di un archivio:

- creazione dell'archivio;
- apertura dell'archivio;
- lettura dei record;
- scrittura dei record;
- inserimento di un nuovo record;
- modifica o aggiornamento di un record;
- cancellazione di un record;
- ordinamento (sort) dei record secondo un certo criterio;
- fusione (merge) tra due o più archivi;
- chiusura dell'archivio;
- cancellazione dell'archivio.

## I/O STANDARD

- input (lettura): è l'operazione che permette di trasferire un dato o un insieme di dati da una unità periferica nella memoria centrale. PERIFERICA → RAM;
- output (scrittura): è l'operazione che permette di trasferire un dato o un insieme di dati dalla memoria centrale ad una unità periferica. RAM → PERIFERICA;
- unità per lo standard input: tastiera;
- unità per lo standard output: monitor.

NB L'hard disk è un'unità periferica di input e output.

## SOFTWARE DI GESTIONE DEI FILE (cenni)

Il modulo del sistema operativo che svolge la funzione di gestione dei file viene chiamato **file system**.

- FAT32
- NTFS



## **CLASSIFICAZIONE MEMORIE (cenni)**

- ▶ Registri della CPU
- ▶ Cache
- ▶ Memoria centrale RAM
- ▶ HD dischi magnetici
- ▶ HD dischi allo stato solido
- ▶ CD, DVD, nastri

## **SISTEMA INFORMATIVO**

Il sistema informativo è l'insieme delle attività svolte da una azienda per organizzare i dati.

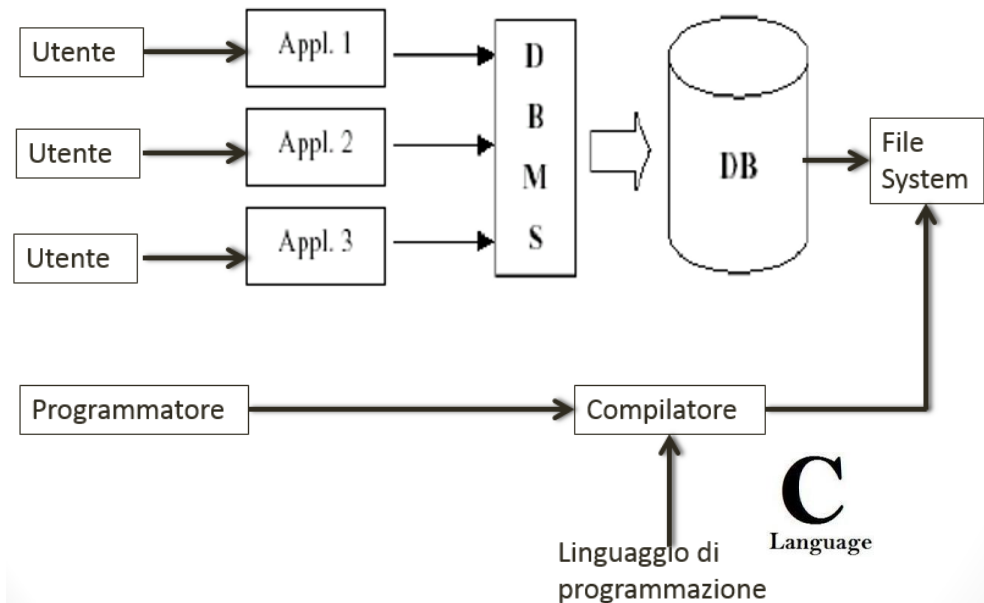
## **SISTEMA INFORMATICO**

È un sistema informatico che utilizza i mezzi informatici al fine di organizzare i dati.



## DALL'ARCHIVIO INFORMATICO TRADIZIONALE ALLA BASE DATI

Per superare i limiti degli archivi gestiti su file con le tecnologie organizzative viste in precedenza, si è affermato a partire dagli 80-90 una nuova tecnologia di salvataggio e organizzazione dei dati facente uso di programmi appositi chiamati DBMS (Database Management System).



Con queste tecniche è possibile raggiungere degli elevati standard di efficienza nel trattamento e nella ricerca dei dati, nella consistenza dell'archivio (i dati devono essere significativi ed effettivamente utilizzabili nelle applicazioni aziendali, i dati devono essere protetti per impedire perdite accidentali cioè occorre garantire un alto grado di integrità dei dati). L'archivio in questo caso prende il nome di database (base di dati). Della base di dati ne parleremo approfonditamente nel corso di questa dispensa.

È molto importante decidere il modo in cui i dati vengono conservati perché ciò determinerà successivamente anche il modo in cui i dati verranno estratti; ad esempio, potrebbe capitare che non preoccupandosi per tempo dell'organizzazione dei dati durante l'archiviazione, si abbiano difficoltà o complicazioni durante la fase del loro recupero.

Mentre una volta gli archivi potevano essere su supporto cartaceo, oggi quasi tutti gli archivi si trovano in forma digitale su supporto elettronico, e ciò rende possibile la loro elaborazione automatizzata. Per esempio, un archivio può avere la forma di un file di dati opportunamente organizzati.

Utilizzare i singoli file per creare archivi di dati può essere pratico per piccole applicazioni, ma in generale, questo approccio, può dar luogo anche a diversi problemi. Per esempio: se una Banca utilizza un file per conservare i dati dei propri clienti e un file per conservare i dati dei relativi conti correnti, bisogna preoccuparsi di mantenere "sincronizzati" i dati tra i due archivi. Se la Banca permette l'aggiornamento degli archivi da parte di più filiali, i problemi aumentano. I database nascono proprio per superare i limiti e gli inevitabili problemi che si potrebbero incontrare usando gli archivi tradizionali. Tra i classici problemi vanno ricordati la *ridondanza* dei dati e l'*inconsistenza* degli archivi. Il primo caso si verifica quando in diversi archivi si trovano memorizzati gli stessi tipi di dati (il problema è quello di dover fare gli aggiornamenti in tutti gli archivi interessati). Il secondo caso si verifica quando questi aggiornamenti non vengono fatti in tutti gli archivi in cui si dovrebbe e quindi nascono incongruenze tra dati nuovi e dati vecchi.

Ricordiamo alcuni termini:

- **Archivio:** oggetto memorizzato in modo permanente, costituito da una sequenza di elementi dello stesso tipo e dimensione (record), ognuno dei quali può essere composto da più campi, di tipo e dimensione diversi tra loro.
- **Database:** insieme di archivi integrati (collegati) secondo un "modello logico" e nel rispetto di determinate regole (vincoli).