



## Introduzione al corso di Basi di Dati

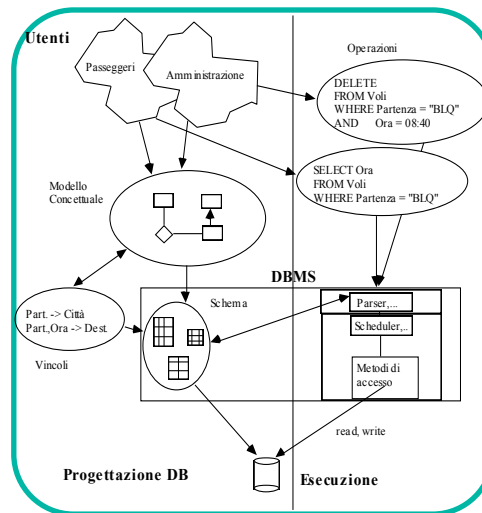


**Dario Maio**

<http://bias.csr.unibo.it/maio>  
dario.maio@unibo.it

### Obiettivi del corso

- **fornire:**
  - le nozioni fondamentali della tecnologia delle basi di dati
  - le linee guida per il progetto di applicazioni DB (e più in generale di un sistema informativo)
- **inquadrare il trend di sviluppo del settore**



## Profilo e ruolo dei S.I.



- ✦ I sistemi informativi (S.I.) affondano le proprie radici sulla correlazione tra **informazione** e **decisione**, e tra **informazione** e **controllo**.
- ✦ **S.I.:** l'insieme di persone, risorse e strumenti che un'organizzazione impiega in modo coordinato con l'obiettivo di acquisire, selezionare e ridistribuire le informazioni utili a individuare strategie di gestione efficienti ed efficaci.
- ✦ Un S.I. deve provvedere alla **raccolta** e alla **classificazione** delle informazioni, da attuarsi con procedure **integrate** e **idonee**, al fine di produrre in **tempo utile** e ai **giusti livelli** le sintesi necessarie per i processi decisionali, nonché per gestire e controllare le attività dell'ente nel suo complesso.

## Tipologia dei S.I. (1)

### + transazionali (TPS: Transaction Processing System)

- + a supporto dei processi operativi e quindi delle decisioni strutturate

### + tattici (MIS: Management Information System)

- + a supporto dei processi gestionali e quindi delle decisioni semi-strutturate

### + di supporto alle decisioni (DSS: Decision Support System)

- + a supporto dei processi direzionali e quindi delle decisioni non-strutturate

## Un esempio di S.I.

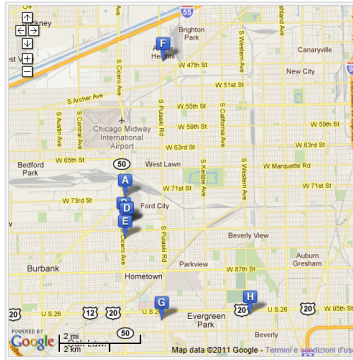
### Store Locator

Advanced Search

By Solution:

Store Name:

Locations near Chicago, IL	iPad	iPhone	iPod	tv	Mac
<b>A Target Store - 0841</b> 7100 S Cicero Ave Chicago, IL 60629					
<b>B Best Buy-311</b> 7600 S Cicero Ave Burbank, IL 60459					
<b>C AT&amp;T</b> 7601 South Cicero Ave Chicago, IL 60652					
<b>D Best Buy-2840</b> 7601 S Cicero Ave Chicago, IL 60652					
<b>E Verizon Wireless</b> 7854 S Cicero Ave Burbank, IL 60459					
<b>F Target Store - 1879</b> 4433 S Pulaski Rd Chicago, IL 60632					
<b>G Target Store - 2087</b> 4120 W 95th St Oak Lawn, IL 60453					
<b>H Sams Club Store - 6349</b> 9400 S Western Ave Evergreen Park, IL 60121					



Find a Solution Professional. Apple Solution Professionals are independent specialists who deliver tailored solutions, training, and product expertise. With experience in Apple and multi-platform technology and industry-specific specialization, Apple Solution Professionals can help you find the right solution for your business.

Buy Online. Open 24-7, the Apple Store online lets you shop anytime from the comfort of your favorite chair. From built-to-order Macs to iPods to printers to cameras, you'll find everything you need and more at the Apple Store online.



Un altro esempio ....

Home FS | Gruppo | Lavora con noi | Carrello | English | Logout | Cerca

**TRENTALIA**  
GRUPPO FERROVIE ITALIANE

Orari e acquisto | Promozioni e offerte | Le Freccie | Carte Fedeltà | In Regione | Aziende | Informazioni per il viaggio | I nostri Treni | Area Clienti

Roma -> Orari e acquisto  
1. scelta del viaggio | 2. tariffe e preferenze | 3. riepilogo e pagamento | 4. conferma acquisto

**AVVATA** | Ordine per: orario di partenza | tempo di percorrenza | numero di cambi | Data: 6/10/2011

Staz. Partenza: Bologna (Tutte Le Stazioni)				Staz. Arrivo: Roma (Tutte Le Stazioni)			
Partenza	Arrivo	Stanza	Staz. Tempo	Categoria	1° classe **	2° classe **	Seleziona
08:23 BOLOGNA	10:45 ROMA TE	02:22	9508 FRECCAROSSA		01.00 €	09.00 €	
09:38 BOLOGNA	10:55 ROMA TE	02:17	9551 FRECCAROSSA		01.00 €	09.00 €	
08:47 BOLOGNA	10:59 ROMA TE	02:03	9451 FRECCARONTO		01.00 €	09.00 €	
08:53 BOLOGNA	11:13 ROMA TE	02:20	9453 FRECCARONTO		01.00 €	09.00 €	
09:12 BOLOGNA	10:24 ROMA TE	01:12	583		02.00 €	08.50 €	

\*\* Il prezzo indicato si intende per un adulto con Biglietto Base.  
Per i treni regionali il prezzo indicato è a Tariffa Intera. Per i treni internazionali è a Tariffa Adulteranda o Pluri, a seconda della relazione.  
Acquisti entro le 16:00 del giorno precedente.

[Stampa](#) | [Dettagli per selezione](#) | [Ridimensiona](#) | [Precedi](#)

© Gruppo Ferrovie 2008 | Contatti | Note Legali | Comunicazioni Legali | Pagine IVA n. 05420151003 | LVR | Condizioni di trasporto | Viaggiare sicuri | FAQ | Call center

Numero unico 199 20 30 60 | Aiuto | Mappa | Accessibilità

Introduzione al corso di Basi di Dati

5



Un altro esempio ...



[SIGN IN](#) | [SIGN UP](#)

[SEARCH](#)

Browse by Title:

1,167 Periodical titles from 296 publishers

Change literature type [\[select a type\]](#)

Switch to [single page view](#)

Powered by **THE ACM GUIDE TO COMPUTING LITERATURE**

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

Abacus - Springer-Verlag New York, Inc.  
Accountability in Research - Gordon and Breach Science Publishers, Inc.  
ACIS International Journal of Computer & Information Science - International Association for Computer & Information Science  
ACM Communications in Computer Algebra - ACM  
ACM Computing Surveys (CSUR) - ACM  
ACM Inroads - ACM  
ACM Journal of Computer Documentation (JCD) - ACM  
ACM Journal on Emerging Technologies in Computing Systems (JETC) - ACM  
ACM Letters on Programming Languages and Systems (LOPLAS) - ACM  
ACM Lisp Bulletin - ACM  
ACM SIGACCESS Accessibility and Computing - ACM  
ACM SIGACT News - ACM  
ACM SIGAda Ada Letters - ACM  
ACM SIGAPL APL Quote Quad - ACM  
ACM SIGAPP Applied Computing Review - ACM  
ACM SIGARCH Computer Architecture News - ACM  
ACM SIGART Bulletin - ACM  
ACM SIGBED Review - ACM  
ACM SIGBIO Newsletter - ACM  
ACM SIGBiomformatics Record - IOS Press  
ACM SIGCAPH Computers and the Physically Handicapped - ACM  
ACM SIGCAS Computers and Society - ACM

Introduzione al corso di Basi di Dati

6



## Tipologia dei S.I. (2)

	Tipo di sistema	Informazioni in input	Elaborazione	Informazioni in output
strategic system	ESS : Executive Support System	Dati aggregati; esterni e interni	Grafici; simulazioni; interattiva	Proiezioni;
	DSS : Decision Support System	Piccoli volumi di dati; modelli analitici	Interattiva; simulazioni, analisi	Rapporti; analisi decisioni;
management system	MIS : Management Information System	Grandi volumi di dati; dati sintetici di transazioni, semplici modelli	Rapporti; semplici modelli; analisi di dettaglio	Sommari e rapporti di errori
	KWS : Knowledge Work System	Specifiche di progetto; basi di conoscenza	Modellazione; simulazioni	Modelli; grafici
knowledge system	OAS : Office Automation System	Documenti; piani di lavoro	Documenti; gestione; pianificazione; comunicazione	Documenti; piani di lavoro; posta
	TPS : Transaction Processing System	Transazioni; eventi	Ordinamenti; liste; aggiornamenti	Rapporti di dettaglio; elenchi; sommari
operational system				

Introduzione al corso di Basi di Dati

7



## S.I. : le solite lamentele



- ✚ abbiamo montagne di dati ma non possiamo accedervi!
- ✚ come è possibile che persone che svolgono lo stesso ruolo presentino risultati sostanzialmente diversi?
- ✚ vogliamo tagliare i dati a fette e a cubetti in ogni modo possibile!
- ✚ mostratemi solo ciò che è importante!
- ✚ tutti sanno che alcuni dati non sono corretti!

R. Kimball, The Data Warehouse Toolkit



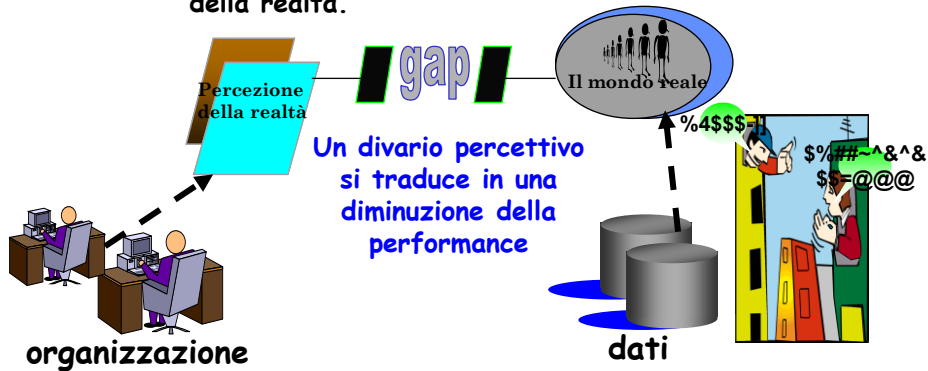
Introduzione al corso di Basi di Dati

8

## *Il divario percettivo*



Un'organizzazione opera  
in base a un modello  
della realtà.

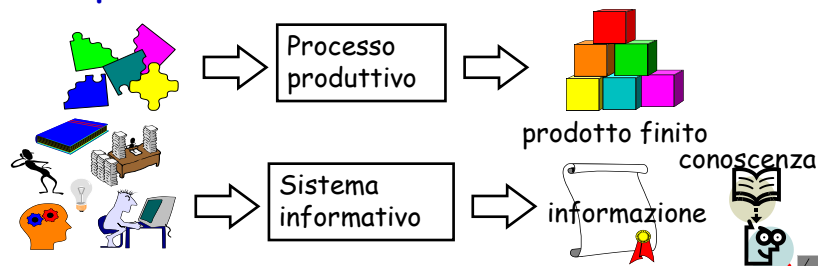


Introduzione al corso di Basi di Dati

9

## *La risorsa informazione*

- L'informazione è un bene, o merce, a valore crescente, richiesto dalla direzione (ma non solo) per pianificare e controllare le attività di un'organizzazione con efficacia.
- I sistemi informativi trasformano dati (e informazioni) in informazioni e conoscenza così come materie prime e semilavorati sono trasformati in prodotti finiti dai sistemi di produzione.



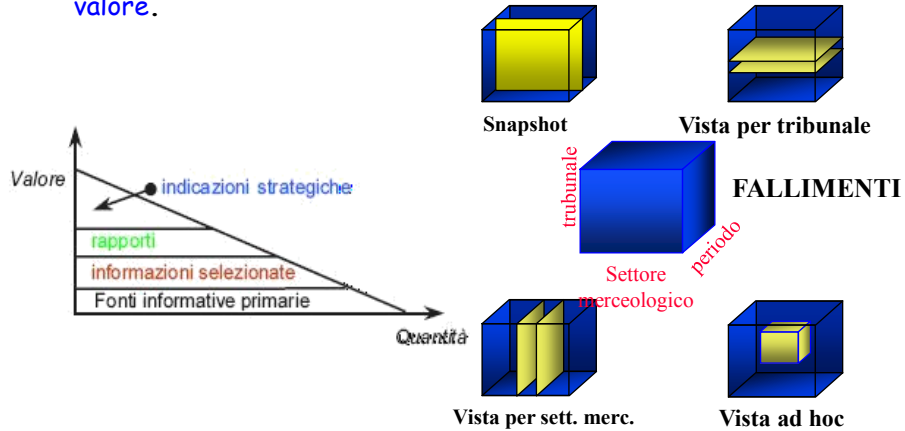
Introduzione al corso di Basi di Dati

10



## Valore dell'informazione

- L'informazione è una risorsa alla stessa stregua del capitale, delle materie prime, degli impianti e delle persone, e come queste ha un costo. Quindi è importante comprenderne il **valore**.

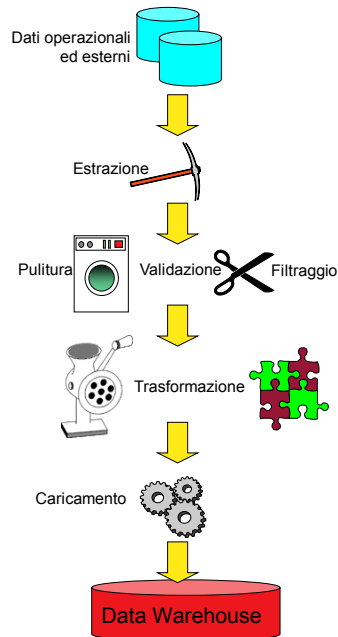


## Data Warehouse

Un DW si appoggia a più fonti di dati eterogenee:

- dati estratti dall'ambiente di produzione, e quindi originariamente archiviati in basi di dati
- dati provenienti da sistemi informativi esterni all'azienda.

Di tutti questi dati il DW restituisce una visione unificata e integrata.



## **Informazioni e processo decisionale**

- Che cosa rende un'informazione utile nel processo decisionale?
- + **Soggettività**: il valore associato a una informazione differisce da individuo a individuo e dipende dal tipo di decisione
- + **Rilevanza**: l'informazione deve essere pertinente alla decisione da prendere
- + **Tempestività**: l'informazione è utile alla decisione solo se è disponibile nel momento decisionale
- + **Accuratezza**: le informazioni devono essere corrette e precise.
- + **Presentazione**: l'informazione deve essere utilizzabile direttamente per la decisione senza ulteriori elaborazioni.
- + **Accessibilità**: le informazioni devono essere disponibili appena necessarie a chi le richiede (considerando le necessarie problematiche di sicurezza e privacy)
- + **Completezza**: il decisore deve avere a disposizione le informazioni necessarie per prendere una decisione corretta

## **Motivazioni per un S.I nel terzo millennio?**



### **1. Globalizzazione**

competizione a livello mondiale,  
gestione e controllo in mercati a larga scala,

....

### **2. Trasformazione dell'economia industriale**

economie basate sulla conoscenza e sull'informazione,  
tempestività delle decisioni, breve durata dei prodotti,

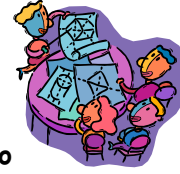
....

### **3. Trasformazione dell'impresa**

decentramento, flessibilità, lavoro collaborativo,

...

## *Impatto della tecnologia dell'informazione (ICT) sull'organizzazione*



Global networks	Indipendenza dal luogo
Enterprise networks	Lavoro cooperativo e di gruppo
Distributed computing	Informazioni disponibili e affidabili
Portable computing	Organizzazione virtuale
Graphical user interfaces	Accessibilità alle conoscenze

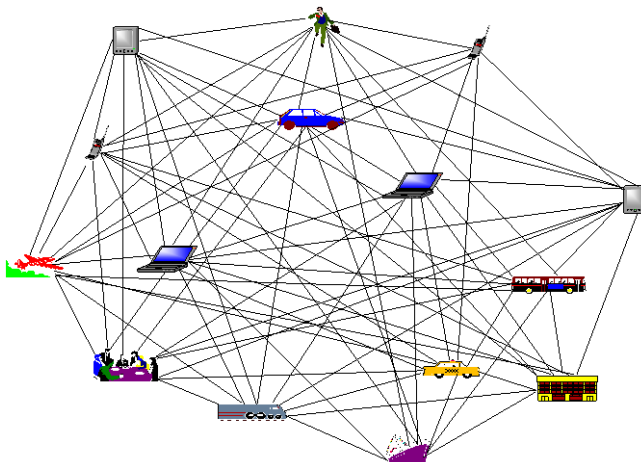


risparmi nei costi delle transazioni, di produzione, di gestione, di coordinamento, di organizzazione

Introduzione al corso di Basi di Dati

15

## *Esempio: Mobile Database System (MDS)*



- Distributed system with mobile connectivity
- Full database system capability
- Complete spatial mobility
- Built on PCS/GSM platform
- Wireless and wired communication capability

Spazio d'informazione interconnesso

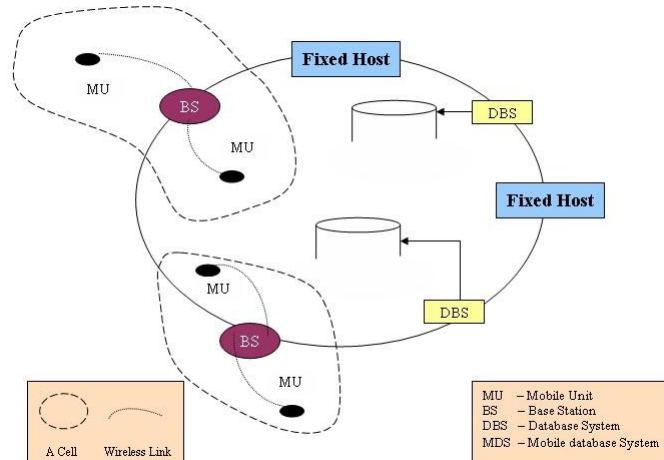
Introduzione al corso di Basi di Dati

16



## Mobile Databases

- Comportano la presenza di tre tipologie di sistemi fixed host, mobile unit, base station.



Introduzione al corso di Basi di Dati

17

## Sistemi Informativi: un processo incrementale non un prodotto



- Visione globale del sistema e del suo ruolo nell'ente
- Quadro concettuale di riferimento
- Adeguate metodologie di progetto e di controllo di qualità
- Appropriate metodologie per il monitoraggio in esercizio
- Esigenze di interoperabilità e scalabilità
- Flessibilità e possibilità "illimitate" di accesso ai dati
- Architettura a elevate prestazioni
- Ridondanza, Sicurezza, Protezione
- Disponibilità continua dei dati
- Interfaccia utente adattativa

Introduzione al corso di Basi di Dati

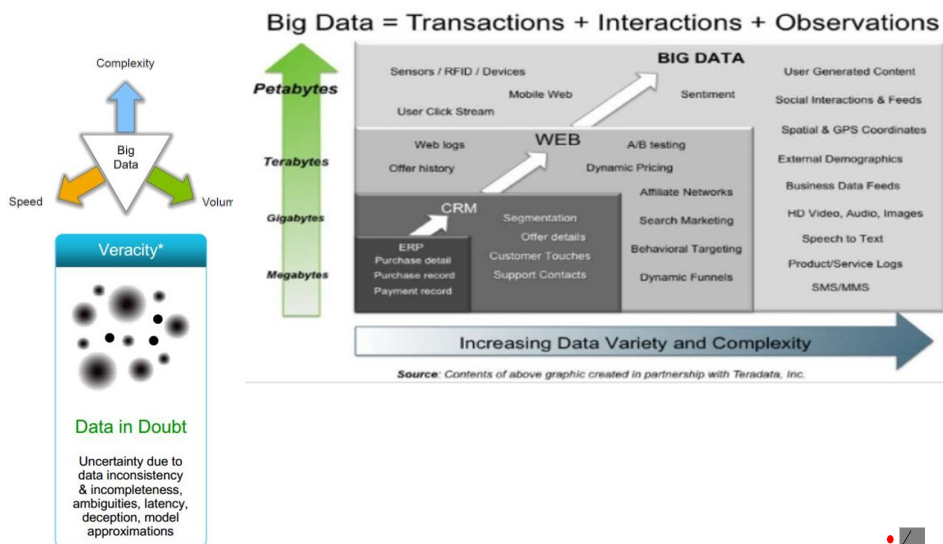
18

## Verso l'integrazione: ERP

**Sistemi di ERP (Enterprise Resource Planning):**  
applicazioni software modulari che permettono di realizzare sistemi integrati e che possono essere personalizzate sulle esigenze e sui processi dell'ente.



## Uno sguardo al futuro: Big Data



## *Chi genera big data?*



**Social media e network**  
(tutti noi generiamo dati)



**Strumentazione scientifica**



**Dispositivi mobili**  
(tracking)



**Reti di sensori sul territorio**  
(misure)

L'innovazione e il progresso non risiedono più nel collezionare dati, ma nella capacità di gestirli, analizzarli, classificarli e scoprire nuove conoscenze in tempo utile e in modo efficace.

Introduzione al corso di Basi di Dati

21

## *Il nuovo paradigma d'interazione*

**Vecchio modello:** Pochi generano dati, tutti gli altri ne fanno uso



**Nuovo modello:** tutti noi generiamo dati e tutti noi ne facciamo uso



Introduzione al corso di Basi di Dati

22

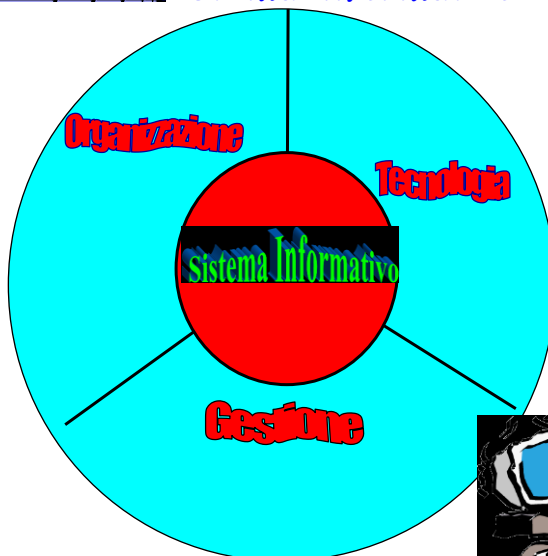
## Ciclo di Vita di un sistema informativo

- definizione strategica (**strategic study**);
- pianificazione (**information system planning**);
- analisi dell'organizzazione (**business analysis**);
- progettazione del sistema (**system design**);
- progettazione esecutiva (**construction design**);
- realizzazione e collaudo in fabbrica (**construction and workbench test**);
- installazione (**installation**);
- collaudo del sistema installato (**test of installed system**);
- esercizio (**operation**);
- evoluzione (**evolution**);
- messa fuori servizio (**phase out**);
- post mortem.

Monitoraggio e  
certificazione di qualità  
nelle varie fasi



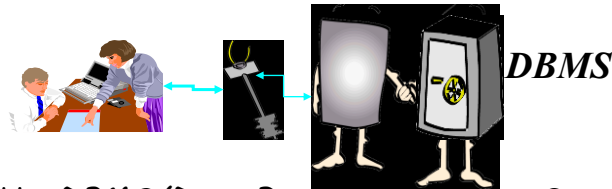
## Sistema informativo e Sistema Informatico



Gran parte della  
attività di un moderno  
sistema informativo  
è svolta facendo ricorso  
alla tecnologia  
dell'informazione



## **Sistema Informativo e DBMS**



✓ Un **DBMS** (Data Base Management System) è un sistema software in grado di gestire efficientemente le informazioni necessarie a un sistema informativo, rappresentandone i dati in forma integrata, e garantendone la persistenza.

✓ Un **RDBMS** è un **DBMS** che offre come modello logico dei dati il **modello relazionale**.

✓ Una **Base di Dati** è una collezione di dati (**DB**). Un **DB** è gestito tramite un **DBMS**.

Introduzione al corso di Basi di Dati

25

## **Uno sguardo al mercato dei DBMS**

Oggi il mercato dei DBMS (il cui fatturato si aggira intorno a  $18 \times 10^9$  US\$) è dominato da un numero ristretto di produttori:

- Oracle
- IBM (DB2 e altri)
- Microsoft (SQL Server)
- Sybase

A questi se ne aggiungono altri del mercato "open-source":

- MySQL
- PostgreSQL
- InterBase

Il fatturato globale del mercato dei DBMS (di cui **l'80% è imputabile ai RDBMS**) cresce di un fattore **> 10%** ogni anno

Introduzione al corso di Basi di Dati

26

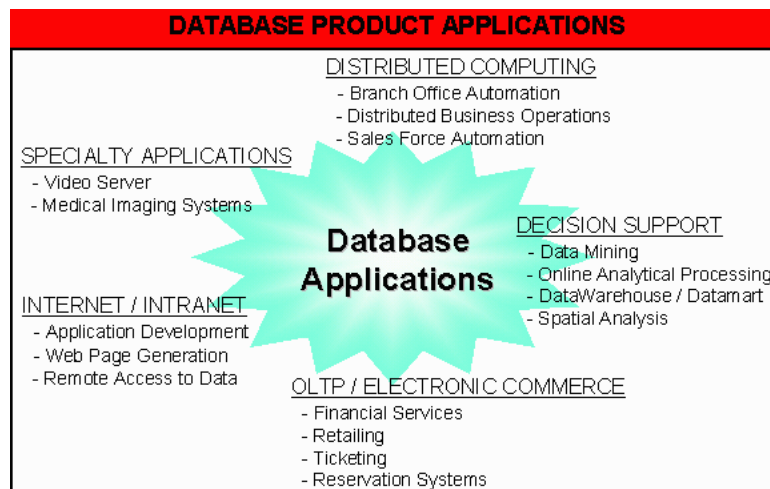
## Perché studiare i DBMS?

- ✦ Dal calcolo all'informazione
  - ✦ sempre vero per **corporate computing**
  - ✦ il web ha reso questo concetto vero anche per **personal computing**
  - ✦ ancora più vero per **scientific computing**
- ✦ Il mercato dei DBMS è esploso negli ultimi anni
  - ✦ **Corporate**: "gestione delle relazioni con il cliente", "gestione della produzione", "logistica", "data warehousing", "data mining", ...
  - ✦ **Scientific**: librerie digitali, progetto Genoma umano, NASA Mission to Planet Earth, sensori, Grid Physics Network, ...
- ✦ Il settore dei DBMS interpreta i fondamenti dell'informatica e le tecnologie ICT rendendo possibile la realizzazione di applicazioni di elevato interesse economico

Introduzione al corso di Basi di Dati

27

## Settori di mercato delle basi di dati



Introduzione al corso di Basi di Dati

28



## Quanto sono grandi i DB?

Norm. Data Volume, All Environments, OLTP							
Company/Organization	Norm. Data Volume (GB)	DBMS	Platform	Architecture	DBMS Vendor	System Vendor	Storage Vendor
UPS	29,329	DB2	z/OS	Federated/SMP	IBM	IBM	IBM
US Bureau of Customs & Border Protection	10,642	CA-Datcom	z/OS	Distributed/SMP	CA	IBM	Hitachi
Elsevier	7,873	Oracle RAC	UNIX	Centralized/Cluster	Oracle	Sun	IBM
Anonymous	5,761	Sybase ASE	UNIX	Centralized/SMP	Sybase	IBM	Hitachi
Turkcell	4,654	Oracle	UNIX	Centralized/SMP	Oracle	HP	EMC
AIM Healthcare Services	4,080	SQL Server	Windows	Centralized/SMP	Microsoft	IBM	EMC
Caixa Econômica Federal	3,733	CA-IDMS	z/OS	Centralized/SMP	CA	IBM	EMC
Amazon.com	2,176	Oracle RAC	Linux	Centralized/Cluster	Oracle	HP	HP
Stadtwerke Munich	1,300	Oracle RAC	Windows	Centralized/Cluster	Oracle	Fujitsu Siemens	EMC
Cellcom	1,279	Oracle	UNIX	Centralized/SMP	Oracle	HP	EMC
Coop	1,231	Oracle	UNIX	Centralized/SMP	Oracle	Sun	Hitachi

Fonte: Winter Corporation, 2005

Introduzione al corso di Basi di Dati

29



## Quante "righe" contengono i DB?

Number of Rows, All Environments, OLTP							
Company/Organization	Number of Rows (Millions)	DBMS	Platform	Architecture	DBMS Vendor	System Vendor	Storage Vendor
UPS	89,621	DB2	z/OS	Federated/SMP	IBM	IBM	IBM
Anonymous	82,639	Sybase ASE	UNIX	Centralized/SMP	Sybase	IBM	Hitachi
Verizon Communications	50,747	SQL Server	Windows	Centralized/SMP	Microsoft	HP	EMC
Anonymous	43,549	DB2	z/OS	Centralized/SMP	IBM	Amdahl	IBM
Experian	35,398	DB2	z/OS	Centralized/SMP	IBM	IBM	EMC
Commander Communications Ltd	31,849	SQL Server	Windows	Centralized/SMP	Microsoft	Dell	EMC
LG Credit Card	30,546	DB2	z/OS	Centralized/Cluster	IBM	IBM	EMC
US Bureau of Customs & Border Protection	24,364	CA-Datcom	z/OS	Distributed/SMP	CA	IBM	Hitachi
Caixa Econômica Federal	20,582	CA-IDMS	z/OS	Centralized/SMP	CA	IBM	EMC
Turkcell	16,855	Oracle	UNIX	Centralized/SMP	Oracle	HP	EMC
AIM Healthcare Services	14,286	SQL Server	Windows	Centralized/SMP	Microsoft	IBM	EMC

Fonte: Winter Corporation, 2005

Introduzione al corso di Basi di Dati

30



## Quanto sono usati i DB2

Circa 300000 al sec !

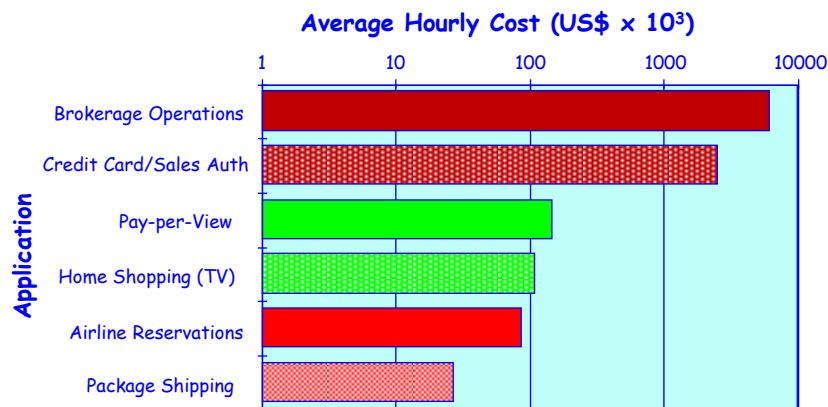
Peak Workload, Mainframe and other OLTP							
Company/Organization	Peak Workload	DBMS	Platform	Architecture	DBMS Vendor	System Vendor	Storage Vendor
UPS	1,134,034,718	DB2	z/OS	Federated/SMP	IBM	IBM	IBM
US Bureau of Customs & Border Protection	340,838,403	CA-Datcom	z/OS	Distributed/SMP	CA	IBM	Hitachi
Experian	202,214,000	DB2	z/OS	Centralized/SMP	IBM	IBM	EMC
State Street Corporation	195,430,140	CA-Datcom	z/OS	Federated/SMP	CA	IBM	EMC
Caixa Econômica Federal	131,847,300	CA-IDMS	z/OS	Centralized/SMP	CA	IBM	EMC
CheckFree Corporation	66,046,711	DB2	z/OS	Centralized/Cluster	IBM	IBM	EMC
LG Credit Card	36,639,038	DB2	z/OS	Centralized/Cluster	IBM	IBM	EMC
Land Registry	6,464,623	DB2	z/OS	Centralized/Cluster	IBM	IBM	IBM
US Department of Treasury/FMS	331,509	CA-Datcom	z/OS	Centralized/SMP	CA	IBM	IBM

**Peak Workload measures the peak SQL statements or database operations per hour**



## L'importanza dei dati

### Financial Impact of System Failure







### 3 punti di vista nel corso

**Utente:** come usare un DB

- modelli dei dati, linguaggi (SQL)



**Progettista:** come progettare un DB e relative applicazioni

- modello E/R, teoria relazionale,
- strumenti per lo sviluppo di applicazioni



**Sistemista (Amministratore):** come gestire un DBMS

- architettura, indici, esecuzione di interrogazioni, transazioni



Introduzione al corso di Basi di Dati

33



### Programma del corso: Parte I

#### ➤ Introduzione ai DBMS:

- gestione di grandi moli di dati, supporto per un modello dei dati, linguaggi di alto livello, gestione delle transazioni, controllo degli accessi, resilienza, ambienti di sviluppo.

#### ➤ Richiami

- sulle caratteristiche costruttive, sui parametri e sull'affidabilità dei dispositivi di memorizzazione dei dati;
- su File System (gestione dello spazio su disco, gestione di richieste concorrenti, buffering di disco, applicazione e file system, organizzazione dei record)

#### ➤ Organizzazioni di file:

- Organizzazione sequenziale, organizzazione ad accesso diretto, fusione e ordinamento di archivi, ISAM, VSAM, Hash file, Indici (B-tree, B\*-tree, B+-tree), Cenni su Indici e strutture multi-dimensionali.



Introduzione al corso di Basi di Dati

34



## Programma del corso: Parte II

### ➤ Modelli dei dati:

- cenni storici sui modelli gerarchico e reticolare. Il modello relazionale (definizioni di base e proprietà). Algebra relazionale operatori di base (selezione, proiezione, unione, differenza, prodotto cartesiano), operatori derivati (intersezione, divisione, join), realizzazione degli operatori, presentazione e valutazione di metodi di join.
- il modello concettuale Entity-Relationship.

### ➤ Il linguaggio SQL

- Non proceduralità e manipolazione set-oriented; i costrutti SQL per la definizione di schemi relazionali e per la manipolazione di relazioni, completezza relazionale, regole di integrità, gestione delle autorizzazioni, gestione delle transazioni, embedded SQL.



## Programma del corso: Parte III

### ➤ Architettura dei DBMS:

- Architettura funzionale di riferimento di un RDBMS, e tipologie di architetture. I principali componenti di un RDBMS: Query Optimizer, Transaction Manager, Scheduler, Recovery Manager, Cache Manager, Storage Manager, Access Methods Manager.

### ➤ Progettazione di DB relazionali

- Introduzione alla progettazione concettuale con schemi E/R. Progettazione logica. Forme normali (1-, 2-, 3-, BCNF). Cenni sulla teoria delle dipendenze e sugli algoritmi di normalizzazione. Cenni sulla Progettazione fisica.

### ➤ Sviluppo di applicazioni DB

- Modalità di accesso ai dati nell'ambiente .NET C# : ADO.NET, Linguaggio Linq

## *Alla fine del corso....*



- \* *Conoscenza degli strumenti necessari per gestire **efficacemente** grandi quantità di dati residenti in **memoria secondaria***
- \* *Capacità di progettare e realizzare applicazioni data base in ambienti relazionali.*
- \* *Conoscenza approfondita dei **metodi di accesso** alle informazioni, conoscenza di base delle strategie di **ottimizzazione** delle operazioni.*

## *Modalità d'esame*

- \* **Discussione di un elaborato di progetto**



- \* **Prova scritta**



- \* **Prova orale**



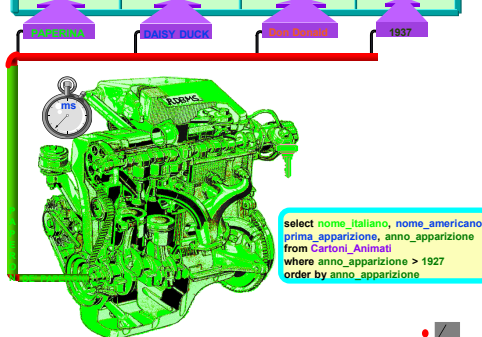
### *Laboratorio*

- SQL Server, Access, Oracle, ...
- DbMain, ERWin, EasyCase, ..
- Microsoft Visual Studio (C#)
- strumenti didattici on-line per esercitazioni e autovalutazione: SQLLab, ExerciseLab

## Testi di riferimento

- **Dispense del docente**
- **D. Maio, S. Rizzi, A. Franco**  
Esercizi di Progettazione di basi di dati  
Esculapio Ed.
- **P. Atzeni, S. Ceri, S. Paraboschi, R. Torlone.**  
Basi di Dati: modelli e linguaggi di interrogazione, McGraw-Hill Italia.
- **P. Atzeni, S. Ceri, S. Paraboschi, R. Torlone.**  
Basi di Dati. Architetture e linee di evoluzione, McGraw-Hill Italia.
- **P. Ciaccia, D. Maio**  
Lezioni di Basi di Dati  
Esculapio Ed.

nome italiano	nome americano	prima apparizione	anno apparizione
TOPOLINO	MICKEY MOUSE	Steamboat Willie	1928
MINNI	MINNIE MOUSE	Steamboat Willie	1928
PLUTO	PLUTO	The Chain Gang	1930
PIPPO	GOOFY	Mickey's Revue	1932
PAPERINO	DONALD DUCK	The Wise Little Hen	1934



Introduzione al corso di Basi di Dati

39

## Bibliografia integrativa

- James R. Groff, Paul N. Weinberg. SQL the Complete Reference 3/E. McGraw-Hill/Osborne Media, 2009.
- J. Gray, A. Reuter. Transaction processing: concepts and techniques. Morgan Kaufmann, 1993.
- R. Ramakrishnan, J. Gehrke. Database Management Systems. McGraw-Hill, 2007.
- R.A. Elmasri, S.B. Navathe. Sistemi di basi di dati - I fondamenti 5/ed. Pearson - Addison Wesley, 2007.

Introduzione al corso di Basi di Dati

40

## *L'essenza del corso in 2 ore (!?)*

### *Le esigenze:*

- *Organizzazione di grandi moli di dati*
- *Strutturazione di dati complessi*
- *Protezione dei dati*
- *Efficienza e semplificazione delle operazioni*
- *Coordinamento degli utenti*
- *Facilità d'uso*

## *Progettazione di una base dati*



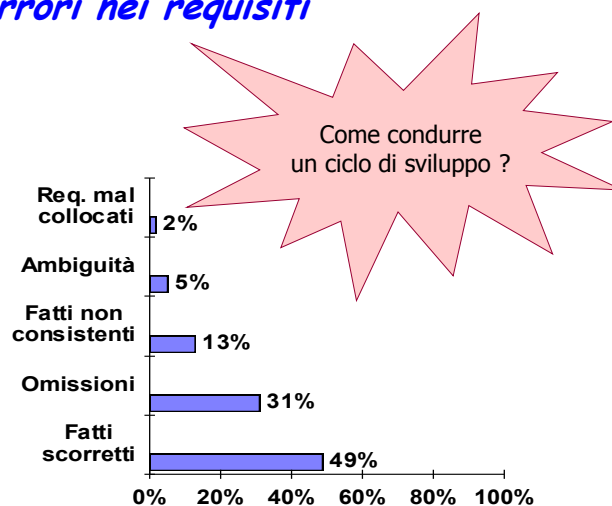
- ✚ La progettazione di una **base di dati** è una delle attività principali del processo di sviluppo di un **sistema informativo**.
- ✚ Di grande rilevanza è l'adozione di **corrette metodologie** per l'analisi e la specificazione dei requisiti.
- ✚ Il processo di analisi è **incrementale** e porta per passi successivi alla stesura di un **insieme di documenti** in grado di rappresentare un modello dell'organizzazione e comunicare, in modo non ambiguo, una **descrizione esauriente, coerente e realizzabile** dei vari aspetti **statici, dinamici e funzionali** di un S.I.

## *Qualità e know how nei sistemi informativi*

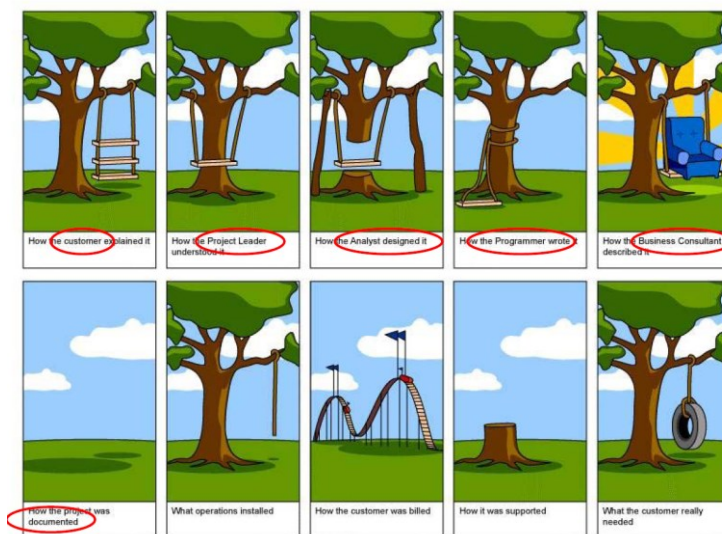
Stadio	Costo relativo per la correzione
Requisiti	0.1 - 0.2
Progettazione	0.5
Codifica	1
Test	2
Accettazione	5
<b>Manutenzione</b>	<b>20</b>

Più tardi viene scoperto un errore nel ciclo di sviluppo, maggiore è il costo di riparazione

## *Tipi di errori nei requisiti*








## *Errori nella conduzione del progetto ...*



Introduzione al corso di Basi di Dati

45

## *Aspetti ontologici da modellare*

-  conoscenza concreta (oggetti e fatti specifici)
-  conoscenza astratta (fatti generali che descrivono la conoscenza concreta)
-  conoscenza procedurale (modalità per modificare la conoscenza concreta)
-  dinamica (modalità per l'evoluzione nel tempo della conoscenza);
-  comunicazione (modalità per accedere alla conoscenza)

Introduzione al corso di Basi di Dati

46



## Tassonomia dei Metodi di Analisi

- ✦ analisi orientata alle **funzioni**
- ✦ analisi orientata agli **oggetti**
- ✦ analisi orientata agli **stati**



- l'orientamento di un metodo verso un aspetto è determinato dall'approccio seguito durante l'analisi e dal diverso accento riposto nella modellazione della realtà.
- la tendenza attuale è integrare modelli di rappresentazione nati per finalità diverse: il linguaggio di modellazione **UML (Unified Modeling Language)** si è affermato come uno standard de facto.



## UML (Unified Modeling Language)

- ✦ **use case diagram** : elenca i casi d'uso e le loro relazioni
- ✦ **class diagram** : descrive la struttura degli oggetti e le loro relazioni
- ✦ **object diagram** : mostra insiemi di oggetti d'interesse e le loro relazioni
- ✦ **sequence diagram** : mostra le interazioni tra oggetti durante scenari di funzionamento privilegiando la sequenza temporale
- ✦ **interaction diagram** : illustra le interazioni tra oggetti durante scenari di funzionamento del sistema
  - ✦ **sequence diagram** : privilegiando la sequenza temporale
  - ✦ **activity diagram** : privilegiando la struttura del sistema stesso
- ✦ **state-transition diagram** : tramite la notazione di Harel descrive gli stati di un oggetto e le sequenze evento-azione-transizione di stato; la stessa notazione consente di descrivere operazioni (activity diagram)
- ✦ **component diagram** : descrive l'architettura fisica del sistema
- ✦ **deployment diagram** : descrive la configurazione hw e l'allocazione dei moduli software





## *Le fasi di progettazione di un S.I. Raccolta e Analisi dei Requisiti*



✚ **Obiettivo:** fornire una **descrizione organica e coerente** dei requisiti informativi illustrando aspetti statici, funzionali e dinamici.



✚ A partire dalle prime interviste, il processo di analisi porta, **per passi successivi di raffinamento**, alla stesura di documenti in grado di rappresentare un modello dell'organizzazione e descrivere i requisiti del sistema informativo da realizzare.



## *Progettazione guidata dai dati*

- ✚ Due aspetti di primaria importanza nella progettazione di un sistema informativo:
  - ✚ progettazione della base dati
  - ✚ progettazione delle applicazioni
- ✚ Il ruolo primario viene molto spesso svolto dai dati, in quanto:
  - ✚ sono strutturalmente più stabili nel tempo
  - ✚ sono condivisi da più applicazioni
- Necessità di una metodologia di applicabilità generale, di facile uso e con supporto di strumenti automatici, in grado di:
  - definire le fasi in cui l'attività di progettazione si articola
  - fornire criteri per scegliere tra diverse alternative
  - rendere disponibili adeguati modelli di rappresentazione
  - fornire supporto per lo sviluppo del software



## Le fasi di progettazione di un S.I. Progettazione concettuale

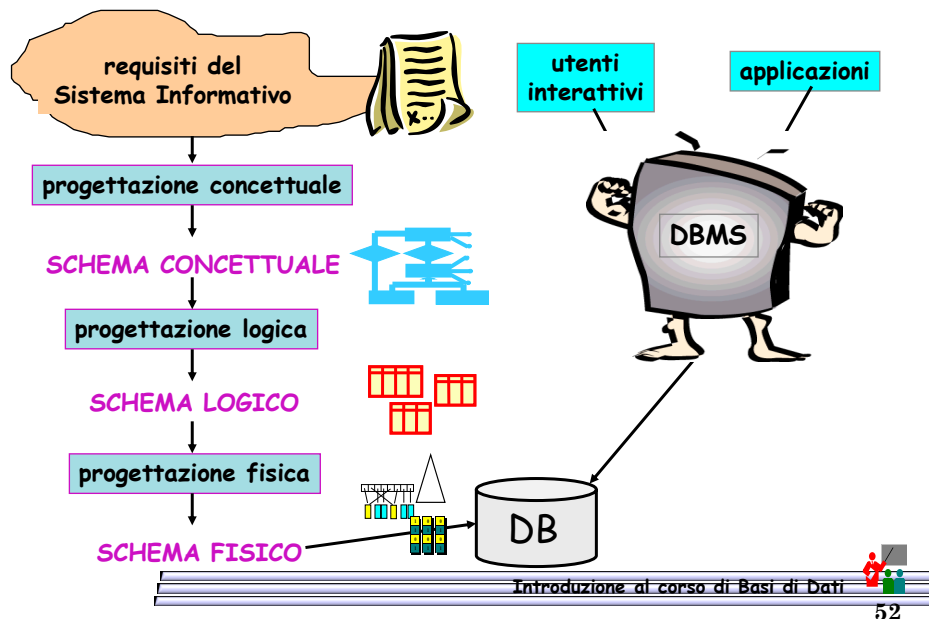
✚ **Obiettivo:** fornire **specifiche indipendenti** dagli strumenti di realizzazione

✚ Un tipico metodo di progettazione concettuale fa uso di:

Dati	<b>Schemi Entity/Relationship</b> per modellare gli aspetti statici
Funzioni	<b>Data Flow Diagram</b> per modellare gli aspetti funzionali
Stati	<b>State Diagram</b> per modellare gli aspetti dinamici

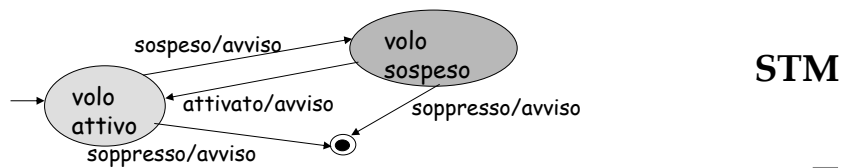
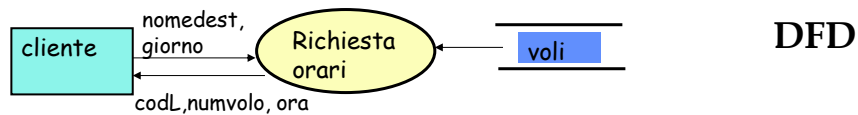
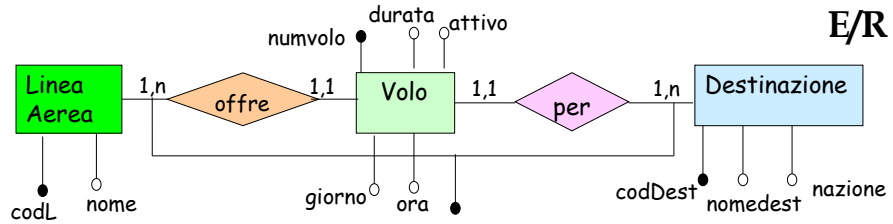


## Progettazione della base di dati





## Esempio di progettazione concettuale dati, funzioni, stati



## Progettazione logica (relazionale)

- **Obiettivo:** descrivere la base dati secondo il modello logico (**relazionale**) adottato dal DBMS (**RDBMS**)

<u>codL</u>	nome
L001	TWA
L002	ALITALIA
....	.....

LINEE

<u>codDest</u>	nomedest	nazione
FCO	ROMA	ITALIA
JFK	NEW YORK	USA
.....	.....	.....

DESTINAZIONI

<u>numvolo</u>	codL	codDest	giorno	ora	durata	attivo
TW056	L001	JFK	LUN	9:00	2	SI
AZ854	L002	FCO	MER	22:30	8	SI
....	.....	.....	.....			

VOLI

Esempio di query SQL

```
select avg(durata)
from VOLI
where codDest='JFK';
```

## Progettazione fisica

- **Obiettivo:** scelta delle strutture di memorizzazione e degli indici (in parte dipendente dal particolare DBMS)

Indice su LINEE.codL (chiave primaria) clustered

Indice su DESTINAZIONI.codDest (chiave primaria) clustered

Indice su DESTINAZIONI.nomedest (chiave secondaria) unclustered

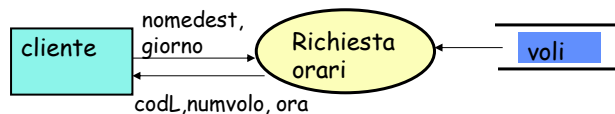
Indice su VOLI.numvolo (chiave primaria) unclustered

Indice su VOLI.codDest (chiave secondaria) clustered

Indice su VOLI.codL (chiave secondaria) unclustered

## Progettazione dell'applicazione

- **Obiettivo:** realizzazione moduli software per le funzioni e per l'interfaccia utente



```
select L.nome, V.numvolo, V.ora
from VOLI V, DESTINAZIONI D, LINEE L
where D.nomedest='New York'
and V.giorno="Lun"
and D.codDest = V.codDest
and L.codL = V.codL;
```

Nome destinazione	<input type="text" value="New York"/>
Giorno	<input type="text" value="Lunedì"/>



## Un background "minimale"

- ✚ Le informazioni gestite da un DBMS convenzionale sono **strutturate**, a differenza di quanto accade, ad esempio, nei sistemi di **Information Retrieval** preposti alla gestione di **documenti non strutturati**.
- ✚ **Campo (dato elementare):**
  - ✚ Unità minima di informazione dotata di significato. È detto anche **attributo**.
- ✚ **Record:**
  - ✚ Raggruppamento di campi relativi a un medesimo "oggetto". I campi sono usati per descrivere proprietà, caratteristiche dell'oggetto.
- ✚ **Chiave (primaria):**
  - ✚ Campo utilizzato per **identificare** un record e, conseguentemente, l'oggetto rappresentato.
- ✚ **Archivio:**
  - ✚ Insieme omogeneo di registrazioni memorizzato su memoria permanente.



## Caratteristiche dei dati

- **Struttura dei record**
  - Quali campi compaiono nel record, quali relazioni sussistono tra tali campi dal punto di vista logico.
- **Tipi di associazioni**
  - Un insieme complesso d'informazioni può essere rappresentato facendo uso di più archivi, e quindi di più tipi di record. È importante conoscere le relazioni precise che esistono tra i diversi archivi.
- **Volatilità**
  - Un archivio è normalmente soggetto a modifiche: più ve ne sono, più l'archivio è "volatile".
- **Espandibilità**
  - Un archivio la cui dimensione varia notevolmente nel tempo può, dipendentemente da come è organizzato, dover essere "riprogettato", passando attraverso una fase detta di "riorganizzazione".
- **Dimensioni**
  - Al crescere della dimensione, aumenta l'esigenza di pensare ad organizzazioni più sofisticate che garantiscano una gestione efficiente dei dati.
- **Periodo di vita**
  - Archivi la cui durata nel tempo è limitata necessitano in minor misura di "cure".

## **Modalità d'uso dei dati**

- **Tipo di operazioni eseguite**
  - Ricerche vs. modifiche, tipi di ricerche, criteri di ordinamento, interattività o meno, ecc.
- **Frequenza di accesso ai dati**
  - È di interesse quando i record di un archivio non sono referenziati in modo uniforme dalle applicazioni.
  - La regola (empirica) detta "80-20" asserisce che l'80% degli accessi riguarda solo il 20% dei dati.
- **Tempi di risposta richiesti**
  - Vi possono essere vincoli (detti di "tempo reale") sul tempo di risposta che intercorre dal momento in cui la richiesta è inoltrata al momento in cui il risultato è prodotto o, più semplicemente, si possono volere dei tempi di risposta "ragionevolmente brevi" per una serie di operazioni interattive.
- **Affidabilità**
  - Come e in quale misura il sistema garantisce l'integrità dei dati a fronte di malfunzionamenti di varia natura.
- **Sicurezza**
  - Come e in quale misura il sistema regola le modalità d'uso dei dati da parte di diversi utenti.

Introduzione al corso di Basi di Dati

59

## **Il database della POOC**

Data	Ora	Cassa	Prodotto	Qtà	Importo
20/12/2005	17.53	21	Panettone GnamGnam 1Kg	2	26000
20/12/2005	17.53	21	Spumante Bollicin 1 lt.	5	60000
20/12/2005	18.01	21	Dentifricio WhiteTeeth	1	3400
20/12/2005	18.02	15	Spumante Bollicin 1 lt.	2	24000
20/12/2005	18.06	3	Caffè BlackMoka 250 gr.	1	4100

- ✓ operazioni "**semplici**" di registrazione (molte)
- ✓ operazioni di "analisi dati" (**complesse**):
  - 1 Importo complessivo delle vendite
  - 2 Importo vendite per giorno, cassa, fascia oraria,...
  - 3 Correlazioni di vendita (chi compra il panettone, spesso compra anche lo spumante)

Introduzione al corso di Basi di Dati

60

## Un esempio: le casse il 20/12/2005

```

openfile(f, `mypath/myfile`);
for i=1 to num_casse
  do somma[i] = 0;
  while not eof(f) do
    read(f,rec);
    if rec.Data = 20/12/2005 then
      somma[rec.Cassa] +=
        rec.Importo
    enddo;
  closefile(f);
  for i=1 to num_casse
    do print(i,somma[i]);
  
```

```

Select  Cassa,sum(Importo)
From    Incassi
Where   Data = 20/12/2005
Group by Cassa;
  
```

non procedurale

Legato a:

- path e file
- formato dei dati
- elaborazione sequenziale

Introduzione al corso di Basi di Dati

61

## Un altro esempio: la cassa 21 nel 2005

```

openfile(f, `mypath/myfile`);
for i=1 to 365 * num_prodotti
  do somma[i] = 0;
  while not eof(f) do
    read(f,rec);
    if rec.Cassa = 21 and
       year(rec.Data) = 2005 then
      i = converti(rec.Data, rec.Prodotto);
      somma[i] += rec.Importo
    enddo;
  closefile(f);
  for i=1 to 365 * num_prodotti
    do print(i,somma[i]);
  
```

```

Select Data,
        Prodotto,sum(Importo)
From    Incassi
Where   Cassa = 21
And     year(Data) = 2005
Group by Data, Prodotto
Order by Data, Prodotto;
  
```

• sort implicito in "converti"

• memoria male utilizzata

• poco chiaro

Introduzione al corso di Basi di Dati

62

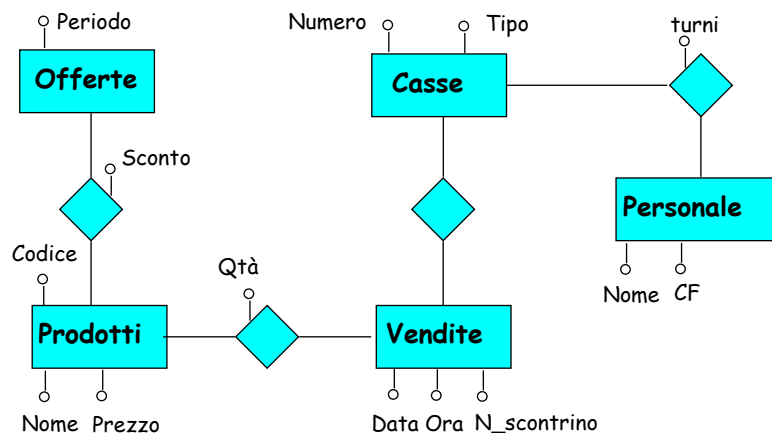
## ... il DB della Pooc si complica

e deve anche gestire:

- *il personale (turni, stipendi, ecc.)*
- *le offerte promozionali*
- *gli ordini ai fornitori*
- ...

*C'è bisogno di strumenti efficaci, chiari e sintetici per rappresentare i dati di interesse e le loro relazioni*

## (parte del) nuovo schema E/R della Pooc



... e poi deve essere completato, raffinato e "tradotto" in tabelle (relazioni)



## *I dati sono "vincolati"*

In ogni situazione reale i dati che si vogliono gestire devono rispettare alcuni **vincoli**

Nel DB della POOC:

- ❑ Ogni cassa ha un numero univoco
- ❑ Ogni persona sta al più a una cassa alla volta
- ❑ Il prezzo di un prodotto in un periodo è costante
- ❑ ...

La conoscenza dei vincoli permette di analizzare la "bontà" di un DB

## ... ancora la tabella Incassi

Data	Ora	Cassa	Prodotto	Qtà	Importo
20/12/2005	17.53	21	Panettone GnamGnam 1Kg	2	26000
20/12/2005	17.53	21	Spumante Bollicin 1 lt.	5	60000
20/12/2005	18.01	21	Dentifricio WhiteTeeth	1	3400
20/12/2005	18.02	15	Spumante Bollicin 1 lt.	2	???
20/12/2005	18.06	3	Caffè BlackMoka 250 gr.	1	4100

- Poiché il prezzo di un prodotto in un dato giorno è sempre lo stesso, il valore nella 4<sup>a</sup> riga non è "nuova" informazione (vale 24000!)

La nostra tabella contiene molti dati ridondanti

## Cosa c'è "sotto"?

- ✓ Un DBMS è un sistema SW (**molto**) complesso ( $\cong 10^5 - 10^6$  LOC) a causa della varietà di funzionalità che mette a disposizione (linguaggi, gestione della memoria, gestione della concorrenza, fault tolerance, ...)
- ✓ come fa un DBMS a eseguire le operazioni richieste?

Fra gli ingredienti di base di un DBMS:

**Metodi di accesso**

**Ottimizzatore**

## Che Cos'è un metodo di accesso?

- È un modo (**interfaccia**) per accedere ai dati, che utilizza una specifica organizzazione "fisica" (**implementazione**).

Es.: un **B<sup>+</sup>-tree** è un albero paginato perfettamente bilanciato che risolve query di intervallo su domini totalmente ordinati in tempo  $O(\lg N + \text{result\_size})$

**index\_scan**(Incassi.Importo,[20000,40000])

	Data	Ora	Cassa	Prodotto	Qtà	Importo
→	20/12/2005	17.53	21	Panettone GnamGnam 1Kg	2	26000
	20/12/2005	17.53	21	Spumante Bollicin 1 lt.	5	60000
	20/12/2005	18.01	21	Dentifricio WhiteTeeth	1	3400
→	20/12/2005	18.02	15	Spumante Bollicin 1 lt.	2	24000
	20/12/2005	18.06	3	Caffè BlackMoka 250 gr.	1	4100

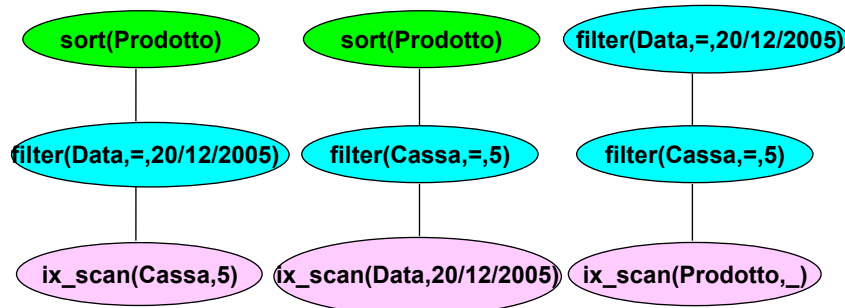
### Cos'è l'ottimizzatore?

- è un modulo del DBMS che si fa carico di determinare il miglior modo per eseguire un'operazione.
- ogni "modo" è detto **piano di accesso**
- sfrutta i metodi di accesso a disposizione e informazioni quantitative (statistiche) sui dati

#### Esempio:

*produci in output, ordinate per prodotto, le vendite della cassa 5 del 20/12/2005*

### 3 piani di accesso



Data	Ora	Cassa	Prodotto	Qtà	Importo
20/12/2005	17.53	21	Panettone GnamGnam 1Kg	2	26000
20/12/2005	17.53	21	Spumante Bollicin 1 lt.	5	60000
20/12/2005	18.01	21	Dentifricio WhiteTeeth	1	3400
20/12/2005	18.02	15	Spumante Bollicin 1 lt.	2	24000
20/12/2005	18.06	3	Caffè BlackMoka 250 gr.	1	4100