

LM87 - Data Mining, statistica sociale ed informatizzazione dei dati, Sistemi per la gestione di basi di dati

Introduction to Relational Databases

Giovanni Giuffrida

Introduzione

- Cenni storici
- Sistemi informativi e sistemi informatici
- Concetto di dato e informazione
- Astrazione dei dati
- Organizzazione relazionale dei dati
- Interrogazione dei dati
- Cenni di Algebra Relazionale
- Database Management System
- Data mining e text mining
- Cenni di algoritmi di data/text mining
- Big Data, Artificial Intelligence e Machine Learning
- Big Data per le scienze sociali

- Slides del corso fornite dal docente
 - Stampare e usarle a lezione per appunti
 - ATTENDERE INDICAZIONI DEL DOCENTE PRIMA DI STAMPARLE
- **Raccomandato:** Big data. Una rivoluzione che trasformerà il nostro modo di vivere e già minaccia la nostra libertà. Viktor Mayer-Schönberger, Kenneth N. Cukier e R. Merlini
- **Raccomandato:** Il computer come macroscopio, Davide Bennato, Franco Angeli editore
- Opzionale: Informatica e Cultura dell'Informazione, Luca Mari, Giacomo Bonanno e Donatella Sciuto
- Opzionale: Introduction to Computational Social Science, Principle and Applications. Claudio Cioffi-Revilla (In inglese)
- Opzionale: ICT & DBMS, Camillo Bella e Gaetano La Rosa, Libreria CULC
- Opzionale: La patente del computer, Federico Tibone, Zanichelli

- Esame scritto
- 36 domande a risposta V/F
- Richiesta prenotazione
- Presentarsi con doc. identità

Introduzione all'informazione e al dato

- Componente (sottosistema) di qualsiasi organizzazione
- Azienda privata, pubblica, info private, famiglia, etc.
- Gestisce (acquisisce, elabora, conserva, produce) le informazioni di interesse
 - ogni organizzazione lo prevede
 - possibilmente non esplicitato nella struttura stessa
 - fondamentalmente di supporto agli altri sottosistemi
 - spesso suddiviso in sottosistemi (gerarchico e/o decentrato)

- Il concetto di “sistema informativo” è indipendente da qualsiasi automazione
- Esistono organizzazioni la cui ragion d'essere è la “gestione di informazioni”
 - Esempio: Anagrafe comunale
- Operano da secoli
- La gestione delle loro informazioni si è evoluta nel tempo...

Ma ciò non altera l'obiettivo dell'organizzazione!

- Porzione automatizzata del sistema informativo
- Insieme di hardware e software per la gestione delle informazioni
- La parte del sistema informativo che gestisce informazioni con tecnologia informatica

Anche prima di essere informatizzati, molti sistemi informativi si sono via via evoluti

- Razionalizzazione delle procedure di accesso ai dati
- Standardizzazione dell'esecuzione delle procedure
- Organizzazione delle informazioni
- Duplicazione degli archivi

- Nelle attività umane, le informazioni vengono gestite (registrate e scambiate) in forme diverse
 - idee informali
 - linguaggio naturale (scritto o parlato, formale o colloquiale, in una lingua o in un'altra)
 - disegni, grafici, schemi
 - numeri e codici
 - audio/video
- E su supporti/dispositivi diversi
- La forma giusta dipende dal contesto e dal supporto utilizzato

La digitalizzazione dell'informazione deve tenere conto di ciò

- Via via è cresciuta l'esigenza di organizzare e codificare le informazioni
- Necessità di identificare *univocamente* le entità
- Ad esempio, nei servizi anagrafici, si è iniziato con registrazioni discorsive e poi
 - nome e cognome
 - estremi anagrafici
 - codice fiscale
- Esigenza di identificare univocamente gli immobili

Storia [[modifica](#) | [modifica wikitesto](#)]

Anticamente le città europee non presentavano numerazione degli edifici. Si poteva individuare un edificio (essenzialmente per scopi fiscali, di cessione o di registrazione delle pigioni) facendo riferimento all'isolato all'interno della [regio](#) (sistema dell'antica Roma), alla parrocchia in cui si trovava ("popolo"), al [borgo](#) (nucleo abitativo nei pressi o fuori da una porta nelle mura), al [quartiere](#) o [sestiere](#), al nome della via o alla vicinanza a un incrocio (detto in alcune zone "canto"). Il sistema generava alcune incertezze, per cui fin dal tardo medioevo i grandi proprietari di fondi immobiliari (cioè essenzialmente le strutture religiose, ospitaliere e le confraternite) iniziarono ad apporre sugli edifici un'[insegna araldica](#) che ricordasse la proprietà e un numero, spesso sul portale, che le identificava con il riferimento negli inventari dell'istituzione. A volte si avevano edifici contigui interessati da questa numerazione, componendo uno schema molto simile a quello odierno, come sul [Pont Notre-Dame](#) a [Parigi](#), almeno dal XV secolo^[1].



Col tempo si rese necessaria una numerazione completa e sistematica degli edifici delle zone urbane, tanto più con le soppressioni degli ordini religiosi, che fece entrare nelle disponibilità degli Stati un amplissimo patrimonio immobiliare da poter gestire in maniera più semplice ed efficiente. La numerazione si basava spesso su un numero unico progressivo, assegnato a ciascun edificio (e non secondo il metodo corrente su ciascun ingresso), partendo da 1 spesso dato al centro del potere politico, fino a numeri a quattro o cinque cifre. Nel Settecento iniziò a venire testato un nuovo sistema in piccole zone urbane, che prevedeva la numerazione individuale di ciascuna strada: nella *New View of London*, del 1708, si ricorda come "a Prescott Street, Goodman's Fields, invece dei cartelli, le case sono distinte da numeri".

Sobborghi parigini vennero numerati negli anni 1720-30 e in quello stesso periodo, per un migliore controllo di polizia, fu numerato il [ghetto ebraico di Praga](#), temuto come sede di "cospiratori" dalle autorità austriache. A metà del secolo XVIII la numerazione degli edifici prese una piega sistematica, anche nei villaggi, soprattutto per scopi militari. Contemporaneamente la numerazione su larga scala interessava le grandi capitali come [Madrid](#), [Londra](#), [Parigi](#) e [Vienna](#), nonché molte altre città^[2].

Il sistema fu applicato a tappeto in tutto il continente dall'inizio del secolo successivo, con la diffusione del [sistema postale](#), mostrando però presto i suoi limiti: difficoltà di trovare gli edifici, schema difficilmente razionalizzabile, facilità di

Cenni storici [\[modifica \]](#) [\[modifica wikitesto \]](#)

Il codice fiscale è stato introdotto con il [Decreto del presidente della Repubblica](#) n. 605 del 29 settembre 1973 per rendere più efficiente l'[amministrazione finanziaria](#) con la creazione dell'[Anagrafe tributaria](#), fortemente voluta dall'allora ministro delle finanze [Bruno Visentini](#). Esso viene attribuito e rilasciato a ciascun cittadino italiano dall'[Agenzia delle entrate](#) e può essere attribuito anche ai cittadini stranieri. Sin dalla riforma a ciascuna persona veniva consegnato un tesserino prima cartaceo poi di plastica con [banda magnetica](#), riportante oltre al codice fiscale anche cognome, nome, sesso, luogo di nascita, provincia di nascita, data di nascita e anno di emissione.

- Nei sistemi informatici (e non solo), le informazioni vengono rappresentate in modo essenziale *attraverso i dati*

- Nei sistemi informatici (e non solo), le informazioni vengono rappresentate in modo essenziale *attraverso i dati*
- Dal Vocabolario della lingua italiana (1987)
 - **dato**: *ciò che è immediatamente presente alla conoscenza, prima di ogni elaborazione; (in informatica) elementi di informazione costituiti da simboli che debbono essere elaborati.*
 - **informazione**: *notizia, dato o elemento che consente di avere conoscenza più o meno esatta di fatti, situazioni, modi di essere.*

- Nei sistemi informatici (e non solo), le informazioni vengono rappresentate in modo essenziale *attraverso i dati*
- Dal Vocabolario della lingua italiana (1987)
 - **dato**: *ciò che è immediatamente presente alla conoscenza, prima di ogni elaborazione; (in informatica) elementi di informazione costituiti da simboli che debbono essere elaborati.*
 - **informazione**: *notizia, dato o elemento che consente di avere conoscenza più o meno esatta di fatti, situazioni, modi di essere.*
- L'Informazione può essere vista come *un'interpretazione semantica dei dati*
 - Dipende dal contesto, dal tempo di osservazione, dall'osservatore, etc.

- Nei sistemi informatici (e non solo), le informazioni vengono rappresentate in modo essenziale *attraverso i dati*
- Dal Vocabolario della lingua italiana (1987)
 - **dato**: *ciò che è immediatamente presente alla conoscenza, prima di ogni elaborazione; (in informatica) elementi di informazione costituiti da simboli che debbono essere elaborati.*
 - **informazione**: *notizia, dato o elemento che consente di avere conoscenza più o meno esatta di fatti, situazioni, modi di essere.*
- L'Informazione può essere vista come *un'interpretazione semantica dei dati*
 - Dipende dal contesto, dal tempo di osservazione, dall'osservatore, etc.

In genere l'informazione è difficile da catturare stabilmente

- I dati devono quindi *essere* interpretati
- **'Mario'** e **'275'** su un foglio di carta sono due dati

- I dati devono quindi *essere* interpretati
- **'Mario'** e **'275'** su un foglio di carta sono due dati
- Se il foglio di carta viene fornito in risposta alla domanda: “A chi mi devo rivolgere per il problema X; qual è il suo interno?”
- I dati sono interpretati per derivare informazione e arricchire conoscenza
- Quindi dipende dal contesto, potrebbe essere:
 - soldi che Mario mi deve
 - giorni lavorativi di Mario
 - numero di followers di Mario
 - Etc. Etc.

- La rappresentazione precisa di forme più ricche di informazione/conoscenza è difficile

- La rappresentazione precisa di forme più ricche di informazione/conoscenza è difficile
- I dati costituiscono spesso una risorsa strategica

- La rappresentazione precisa di forme più ricche di informazione/conoscenza è difficile
- I dati costituiscono spesso una risorsa strategica
- Più stabili nel tempo di altre componenti quali processi, tecnologie, ruoli umani, etc.

Concetti di Database Management Systems (DBMS)

Insieme organizzato di dati utilizzati per il supporto allo svolgimento dell'attività di interesse

- Tecnologicamente parlando: Insieme di dati gestito da un sistema software dedicato alla “Gestione di basi di dati”

Data Base Management System (DBMS)

Applicazione software in grado di gestire collezioni di dati che siano:

- **Grandi:** da non poter essere gestiti facilmente con strumenti familiari
- **Persistenti:** con un periodo di vita indipendente dalle singole esecuzioni dei programmi che le utilizzano
- **Affidabili:** resistenza a malfunzionamenti hardware e software, blackout, etc.
- **Condivise:** accessibili da più utenti contemporaneamente anche da località geografiche diverse
- **Private:** con una disciplina e un controllo degli accessi

- Un'organizzazione (specie se grande) è divisa in settori o comunque svolge diverse attività
- A ciascun settore corrisponde un (sotto-)sistema informativo
- Possono esistere sovrapposizioni fra i dati di interesse dei vari settori

Possibili problemi nella condivisione dei dati

- **Ridondanza:** informazioni duplicate
- **Rischio di incoerenza:** le versioni possono non coincidere

- **Ridondanza:** informazioni duplicate
- **Rischio di incoerenza:** le versioni possono non coincidere
- *L'incoerenza nei dati è un problema frequente nelle grandi organizzazioni, e spesso molto complicato da risolvere*

- **Ridondanza:** informazioni duplicate
- **Rischio di incoerenza:** le versioni possono non coincidere
- *L'incoerenza nei dati è un problema frequente nelle grandi organizzazioni, e spesso molto complicato da risolvere*

Il DBMS è una risorsa integrata e condivisa fra i vari settori che aiuta significativamente a risolvere il problema dell'incoerenza del dato

- Utenti possono parallelamente accedere agli stessi dati
- In lettura e/o scrittura
- Quindi abbiamo bisogno di:
 - Controllo degli accessi
 - Autorizzazione accesso in Lettura/Scrittura
 - Autorizzazione per utente a porzioni di dati

- Necessità di gestione del dato in “concorrenza”
- Controllo della concorrenza tramite il concetto di “Transazione”
- **Transazione:** attività frequente/predefinita, con poche eccezioni, prevista a priori
- Esempi:
 - versamento in banca
 - prelievo da bancomat
 - prenotazione aerea
 - acquisto biglietti per eventi
 - etc.

Sequenza indivisibile di operazioni

Sequenza indivisibile di operazioni

- **Atomicità:** O vengono eseguite tutte le operazioni o nessuna
- I DBMS sono in grado di garantire l'atomicità delle transazioni
- Costrutti per marcare l'inizio e la fine della transazione

Modello dei dati



Modello concettuale



Modello logico



Modello fisico



- Modellano i “concetti” del mondo reale
- Indipendenti da ogni implementazione fisica
- Utilizzati nelle fasi preliminari di progettazione
- Non richiedono competenze tecnologiche
- Utilizzati dagli “analisti” per “modellare” la realtà
- Esistono vari formalismi: UML, OMT, IE, ...

Entity-Relationship (ER) conceptual model

- Più antico e noto modello concettuale
- Introdotto da Peter Chen, 1976
- **ER**: Entità e relazioni tra esse
- Si *descrivono* le parti di business necessarie
- Modello grafico
- Molto intuitivo
- Precede la concettualizzazione logica
- Non descrive i processi

Entity-Relationship (ER) conceptual model

- Più antico e noto modello concettuale
- Introdotto da Peter Chen, 1976
- **ER**: Entità e relazioni tra esse
- Si *descrivono* le parti di business necessarie
- Modello grafico
- Molto intuitivo
- Precede la concettualizzazione logica
- Non descrive i processi

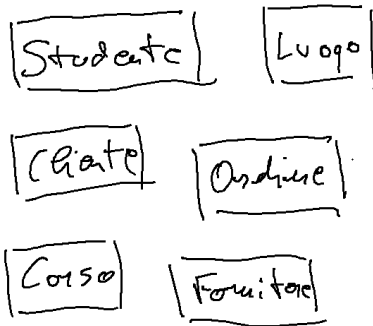
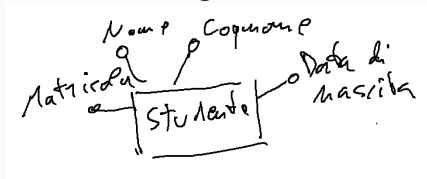
"The entity-relationship model adopts the more natural view that the real world consists of entities and relationships. It incorporates some of the important semantic information about the real world." [1]

1. *Individuare le entità*
2. *Individuare le relazioni tra esse*

Entità

- "Cose" del mondo reale
- Representate graficamente da un rettangolo
- In genere si modellano solo quelle essenziali
- Semplici da individuare

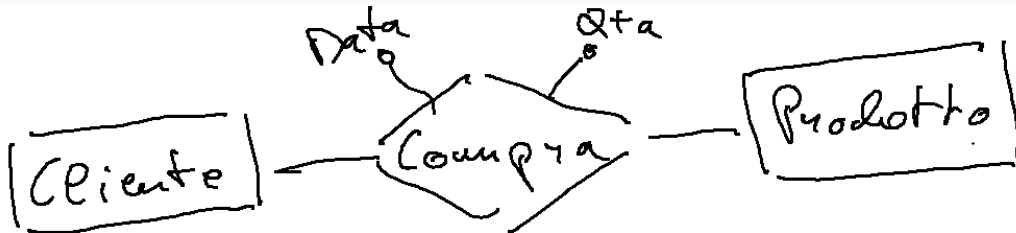
Possono avere degli attributi



- Indicano *relazioni* tra le entità
- Indicate con un rombo
- Recentemente ridenominate come “associazioni”
- Possono avere degli attributi
- Indicate con verbo o sostantivo

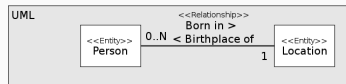
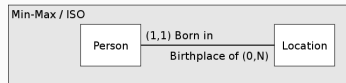
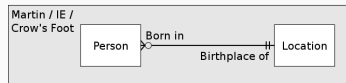
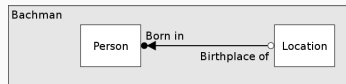
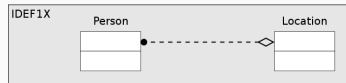
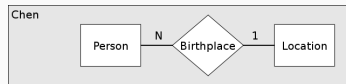
Relazioni/Associazioni

- Indicano *relazioni* tra le entità
- Indicate con un rombo
- Recentemente ridenominate come “associazioni”
- Possono avere degli attributi
- Indicate con verbo o sostantivo



Cardinalità delle relazioni

- Specifica la modalità di "partecipazione" di un'entità alla relazione
- Vari tipi
- Essenziali per modellare opportunamente
- Varie rappresentazioni grafiche



- **One-to-one**: obbligatoria, singola
- **One-to-many**: obbligatoria, multipla
- **Zero-to-one**: opzionale, singola
- **Zero-to-many**: opzionale, multipla

• STUDENTI — Esame — CORSI

- **One-to-one**: obbligatoria, singola
- **One-to-many**: obbligatoria, multipla
- **Zero-to-one**: opzionale, singola
- **Zero-to-many**: opzionale, multipla

- STUDENTI — Esame — CORSI
- IMPIEGATI — Lavora — PROGETTI

- **One-to-one**: obbligatoria, singola
- **One-to-many**: obbligatoria, multipla
- **Zero-to-one**: opzionale, singola
- **Zero-to-many**: opzional, multipla

- **One-to-one**: obbligatoria, singola
- **One-to-many**: obbligatoria, multipla
- **Zero-to-one**: opzionale, singola
- **Zero-to-many**: opzionale, multipla

- STUDENTI — Esame — CORSI
- IMPIEGATI — Lavora — PROGETTI
- IMPIEGATI — Dirige — PROGETTI

- **One-to-one**: obbligatoria, singola
 - **One-to-many**: obbligatoria, multipla
 - **Zero-to-one**: opzionale, singola
 - **Zero-to-many**: opzionale, multipla
- STUDENTI — Esame — CORSI
 - IMPIEGATI — Lavora — PROGETTI
 - IMPIEGATI — Dirige — PROGETTI
 - PERSONE — Possiede — PASSAPORTI

- **One-to-one**: obbligatoria, singola
 - **One-to-many**: obbligatoria, multipla
 - **Zero-to-one**: opzionale, singola
 - **Zero-to-many**: opzionale, multipla
- STUDENTI — Esame — CORSI
 - IMPIEGATI — Lavora — PROGETTI
 - IMPIEGATI — Dirige — PROGETTI
 - PERSONE — Possiede — PASSAPORTI
 - CLIENTI — Acquista — PRODOTTI

- **One-to-one**: obbligatoria, singola
 - **One-to-many**: obbligatoria, multipla
 - **Zero-to-one**: opzionale, singola
 - **Zero-to-many**: opzionale, multipla
- STUDENTI — Esame — CORSI
 - IMPIEGATI — Lavora — PROGETTI
 - IMPIEGATI — Dirige — PROGETTI
 - PERSONE — Possiede — PASSAPORTI
 - CLIENTI — Acquista — PRODOTTI
 - MAGAZZINI — Contiene — PRODOTTI

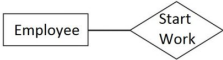
- **One-to-one**: obbligatoria, singola
 - **One-to-many**: obbligatoria, multipla
 - **Zero-to-one**: opzionale, singola
 - **Zero-to-many**: opzional, multipla
- STUDENTI — Esame — CORSI
 - IMPIEGATI — Lavora — PROGETTI
 - IMPIEGATI — Dirige — PROGETTI
 - PERSONE — Possiede — PASSAPORTI
 - CLIENTI — Acquista — PRODOTTI
 - MAGAZZINI — Contiene — PRODOTTI
 - SCONTRINI — Contiene — PRODOTTI

- **One-to-one**: obbligatoria, singola
 - **One-to-many**: obbligatoria, multipla
 - **Zero-to-one**: opzionale, singola
 - **Zero-to-many**: opzionale, multipla
- STUDENTI — Esame — CORSI
 - IMPIEGATI — Lavora — PROGETTI
 - IMPIEGATI — Dirige — PROGETTI
 - PERSONE — Possiede — PASSAPORTI
 - CLIENTI — Acquista — PRODOTTI
 - MAGAZZINI — Contiene — PRODOTTI
 - SCONTRINI — Contiene — PRODOTTI
 - CLIENTI — Possiede — CARTE FEDELTA

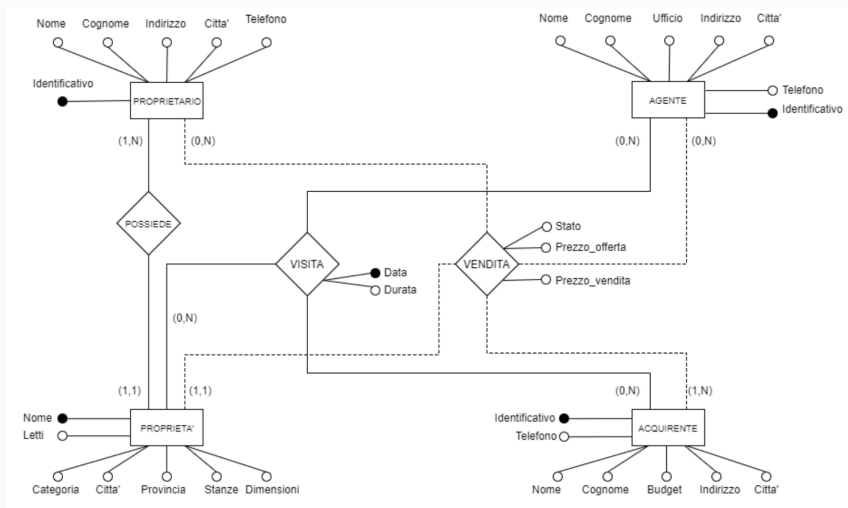
- **One-to-one**: obbligatoria, singola
 - **One-to-many**: obbligatoria, multipla
 - **Zero-to-one**: opzionale, singola
 - **Zero-to-many**: opzional, multipla
- STUDENTI — Esame — CORSI
 - IMPIEGATI — Lavora — PROGETTI
 - IMPIEGATI — Dirige — PROGETTI
 - PERSONE — Possiede — PASSAPORTI
 - CLIENTI — Acquista — PRODOTTI
 - MAGAZZINI — Contiene — PRODOTTI
 - SCONTRINI — Contiene — PRODOTTI
 - CLIENTI — Possiede — CARTE FEDELTA
 - PERSONE — Sposa — PERSONE

- **One-to-one**: obbligatoria, singola
 - **One-to-many**: obbligatoria, multipla
 - **Zero-to-one**: opzionale, singola
 - **Zero-to-many**: opzional, multipla
- STUDENTI — Esame — CORSI
 - IMPIEGATI — Lavora — PROGETTI
 - IMPIEGATI — Dirige — PROGETTI
 - PERSONE — Possiede — PASSAPORTI
 - CLIENTI — Acquista — PRODOTTI
 - MAGAZZINI — Contiene — PRODOTTI
 - SCONTRINI — Contiene — PRODOTTI
 - CLIENTI — Possiede — CARTE FEDELTA
 - PERSONE — Sposa — PERSONE
 - PERSONE — Figlio — PERSONE

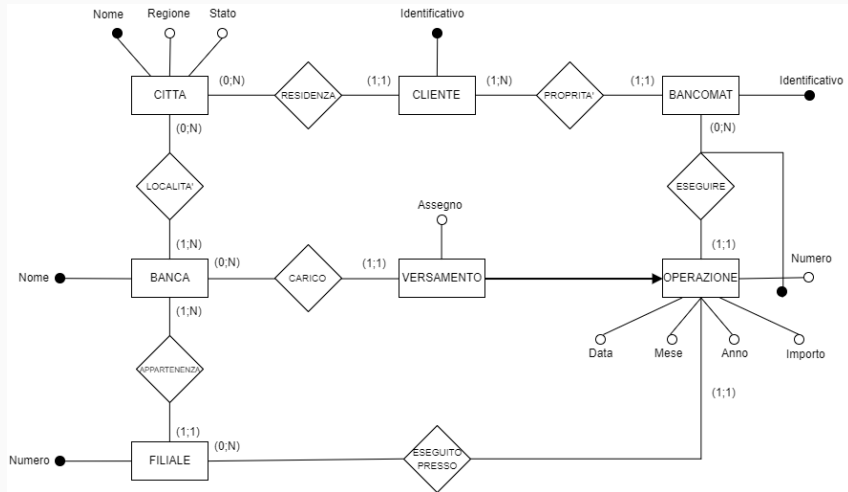
ER symbols

SYMBOL	NAME	MEANING	EXAMPLE
	Rectangle	Entity Set	
	Oval	Attribute	
	Diamond	Relationship	
	Line	Links: Attribute to Entity	
		Entity Set to Relationship	
		Attribute to Relationship	

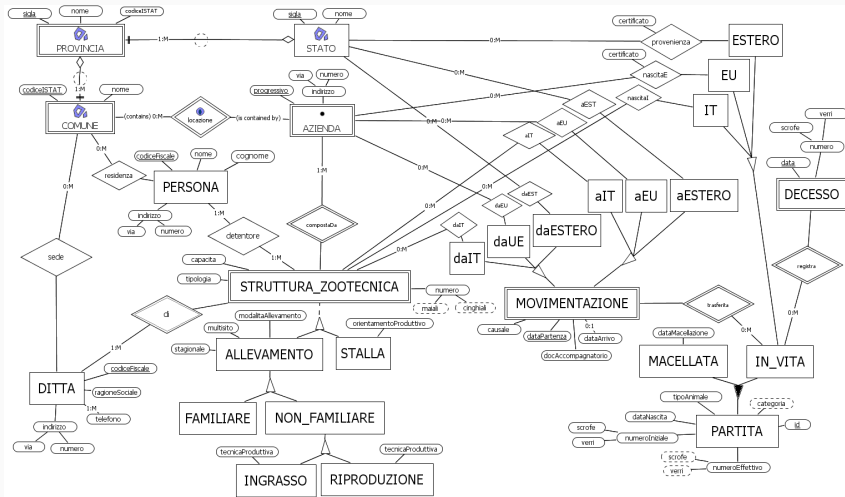
ER Example



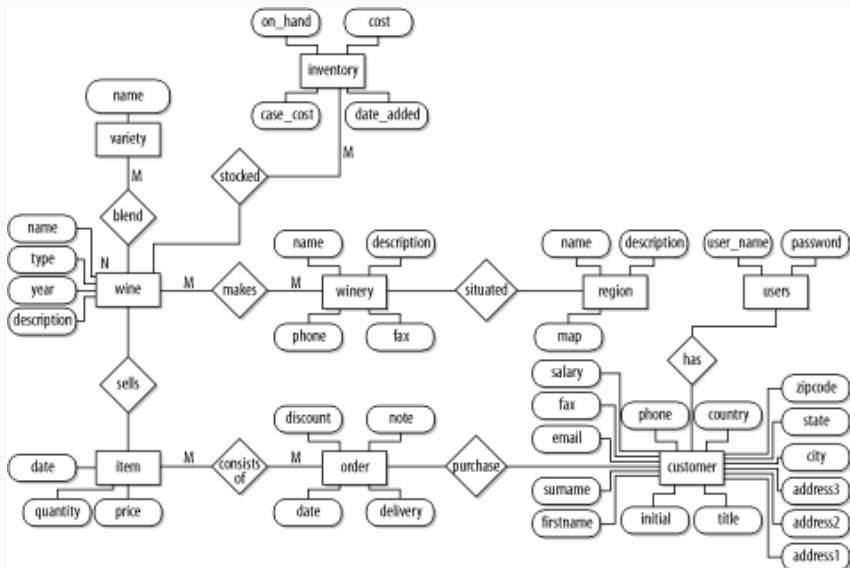
ER Example



ER Example



ER Example



- Il *modello concettuale* modella la realtà
- E' totalmente indipendente dalla realizzazione fisica della base di dati
- Chi crea il modello concettuale non ha bisogno di conoscere nulla di DBMS

Dal modello concettuale al modello logico

- Il *modello concettuale* modella la realtà
- E' totalmente indipendente dalla realizzazione fisica della base di dati
- Chi crea il modello concettuale non ha bisogno di conoscere nulla di DBMS
- Il *modello logico* è “più vicino” all'implementazione fisica, ma è sempre indipendente da essa
- Il modello logico per eccellenza è il **modello relazionale**
- Il modello ER si “mappa” facilmente sul modello relazionale
- Software esistenti per la traduzione automatica

Il Modello Relazionale di Codd

A Relational Model of Data for Large Shared Data Banks

E. F. CODD

IBM Research Laboratory, San Jose, California

Future users of large data banks must be protected from having to know how the data is organized in the machine (the internal representation). A prompting service which supplies such information is not a satisfactory solution. Activities of users at terminals and most application programs should remain unaffected when the internal representation of data is changed and even when some aspects of the external representation are changed. Changes in data representation will often be needed as a result of changes in query, update, and report traffic and natural growth in the types of stored information.

Existing noninferential, formatted data systems provide users with tree-structured files or slightly more general network models of the data. In Section 1, inadequacies of these models are discussed. A model based on n -ary relations, a normal form for data base relations, and the concept of a universal data sublanguage are introduced. In Section 2, certain operations on relations (other than logical inference) are discussed and applied to the problems of redundancy and consistency

The relational view (or model) of data described in Section 1 appears to be superior in several respects to the graph or network model [3, 4] presently in vogue for non-inferential systems. It provides a means of describing data with its natural structure only—that is, without superimposing any additional structure for machine representation purposes. Accordingly, it provides a basis for a high level data language which will yield maximal independence between programs on the one hand and machine representation and organization of data on the other.

A further advantage of the relational view is that it forms a sound basis for treating derivability, redundancy, and consistency of relations—these are discussed in Section 2. The network model, on the other hand, has spawned a number of confusions, not the least of which is mistaking the derivation of connections for the derivation of relations (see remarks in Section 2 on the “connection trap”).

Finally, the relational view permits a clearer evaluation of the scope and logical limitations of present formatted data systems, and also the relative merits (from a logical standpoint) of competing representations of data within a single system. Examples of this clearer perspective are cited in various parts of this paper. Implementations of systems to support the relational model are not discussed.

1.2. DATA DEPENDENCIES IN PRESENT SYSTEMS

The provision of data description tables in recently de-

- L'elemento di base è la **relazione**: una semplice tabella bidimensionale
- Sia le entità che le relazioni del modello ER si mappano su un solo concetto di relazione (confusione in Italiano)

Nome	Matricola	Residenza	Nascita	Età	Sesso
Mario	456	Milano	Roma	22	M
Teo	678	Bergamo	Firenze	30	M
Anna	323	Firenze	Catania	19	F
Giovanni	213	Milano	Padova	22	M
Francesca	112	Firenze	Catania	20	F

- **Relazione:** La tabella
- **Attributi/Colonne:** Nomi delle colonne
- **Tuple/Righe/Record:** Righe della relazione
- **Query:** Interrogazione

Nome	Matricola	Residenza	Nascita	Età	Sesso
Mario	456	Milano	Roma	22	M
Teo	678	Bergamo	Firenze	30	M
Anna	323	Firenze	Catania	19	F
Giovanni	213	Milano	Padova	22	M
Francesca	112	Firenze	Catania	20	F

- Forte connotazione matematica
- “Closure”: Le operazioni sulle relazioni producono una relazione
- Possiamo quindi “provare” matematicamente la validità
- Basato sui valori
- Intuitivo per gli umani, complesso da un punto di vista implementativo

Le relazioni

- Una semplice tabella rappresenta una *relazione*
- La relazione è formata da uno **schema** e dal **corpo**
- I valori di ogni colonna sono fra loro omogenei
- Le intestazioni delle colonne sono diverse tra loro
- Le righe sono diverse fra loro
- Relazione vista come insieme
- L'ordinamento delle righe e delle colonne è irrilevante

Casa	Ospite	G Casa	G Ospite	Data
Lazio	Milan	3	1	14/05/16
Milan	Roma	0	2	21/05/16
Juve	Napoli	1	1	21/05/16
Roma	Inter	2	0	14/05/16

Studenti

Nome	Matricola	Residenza	Nascita	Età	Sesso
Mario	456	Milano	Roma	22	M
Teo	678	Bergamo	Firenze	30	M
Anna	323	Firenze	Catania	19	F
Giovanni	213	Milano	Padova	22	M
Francesca	112	Firenze	Catania	20	F

Corsi

Denominazione	Docente
Informatica	Turing
Matematica	Turing
Diritto	Roosvelt
Storia	Darwin
Letteratura	Roosvelt

Esami

Matricola	Corso	Anno	Voto	Lode
213	Informatica	2017	22	
456	Diritto	2017	30	Si
213	Diritto	2018	18	
112	Matematica	2017	30	
323	Informatica	2017	24	
456	Storia	2018	30	Si
323	Fisica	2018	30	

Informazione incompleta

- Il modello relazionale impone ai dati una struttura rigida
 - le informazioni sono rappresentate per mezzo di ennuple
 - le ennuple devono essere omogenee con lo schema
- I dati disponibili possono non corrispondere al formato previsto

First Name	Second Name	Lastname
Winston	Delano	Churchill
Charles		De Gaulle
Franklin		Roosevelt
Josip		Stalin

- *Sbagliando*, spesso, si usano valori “non utilizzati” del dominio:
 - Stringa vuota, -1 per età non conosciuta, 99 per numero di figli, etc.
- Potrebbero non esistere valori “non utilizzati”
- Valori “non utilizzati” potrebbero diventare significativi nel tempo
- In fase di utilizzo bisogna “ricordarsi” tutte le eccezioni

- Il modello relazionale adotta una soluzione molto elegante: **NULL**
- Non appartiene a nessun *dominio*
- Denota l'assenza di informazione
- Vari casi per assenza di dato:
 - valore inesistente
 - valore sconosciuto
 - valore senza informazione
- I DBMS non distinguono i tipi di valore NULL

Vincoli di integrità

- Dati sintatticamente corretti, ma semanticamente errati

Studenti

Nome	Matricola	Residenza	Nascita	Età	Sesso
Anna	323	Firenze	Catania	19	F
Giovanni	213	Milano	Padova	176	M
Francesca	112	Firenze	Catania	20	F

Esami

Matricola	Corso	Anno	Voto	Lode
213	Informatica	2017	22	Si
987	Diritto	2017	30	
213	Diritto	2018	32	
112	Matematica	2057	30	Si
323	Informatica	2017	24	

- Un vincolo che deve essere sempre soddisfatto
- Tipo di vincolo:
 - **intrarelazionale**: definito sui valori della tupla stessa, es: $18 \leq Voto \leq 30$
 - **interrelazionali**: verifica coincidenza tra valori in relazioni diverse
- Da verificare ad ogni cambio dell'istanza della base di dati: ci pensa il DBMS
- Funzione booleana (predicato)

- Definita su una singola relazione
- Insieme di attributi che identificano univocamente una riga di una tabella
- In genere ogni entità (i.e., relazione) ha una chiave
- Meccanismo largamente utilizzato nei DBMS
- Rappresenta un vincolo di integrità: Il DBMS garantisce la validità della chiave

- Normalmente indicata nello schema con sottolineatura

Nome	<u>Matricola</u>	Residenza	Nascita	Età	Sesso
Mario	456	Milano	Roma	22	M
Teo	678	Bergamo	Firenze	30	M
Anna	323	Firenze	Catania	19	F
Giovanni	213	Milano	Padova	26	M
Francesca	112	Firenze	Catania	20	F

- Esiste la chiave in questa tabella?

Matricola	Corso	Anno	Voto	Lode
213	Informatica	2017	22	Si
456	Diritto	2017	30	
213	Diritto	2018	22	
112	Matematica	2017	30	
323	Informatica	2017	24	Si
456	Storia	2018	30	
323	Fisica	2018	30	

Esempio di chiave

- Esiste la chiave in questa tabella?

<u>Matricola</u>	<u>Corso</u>	Anno	Voto	Lode
213	Informatica	2017	22	Si
456	Diritto	2017	30	
213	Diritto	2018	22	
112	Matematica	2017	30	
323	Informatica	2017	24	Si
456	Storia	2018	30	
323	Fisica	2018	30	

- Insieme dei campi “univoci”: **Matricola** e **Corso** nel nostro caso

Studenti

Nome	Matricola	Residenza	Nascita	Età	Sesso
Mario	456	Milano	Roma	22	M
Teo	678	Bergamo	Firenze	30	M
Anna	323	Firenze	Catania	19	F
Giovanni	213	Milano	Padova	22	M
Francesca	112	Firenze	Catania	20	F

Corsi

Denominazione	Docente
Informatica	Turing
Matematica	Turing
Diritto	Roosvelt
Storia	Darwin
Letteratura	Roosvelt

Esami

Matricola	Corso	Anno	Voto	Lode
213	Informatica	2017	22	
456	Diritto	2017	30	Si
213	Diritto	2018	18	
112	Matematica	2017	30	
323	Informatica	2017	24	
456	Storia	2018	30	Si
323	Fisica	2018	30	

Thanks