



UNIVERSITÀ DEGLI STUDI DI SALERNO
DIPARTIMENTO DI INFORMATICA



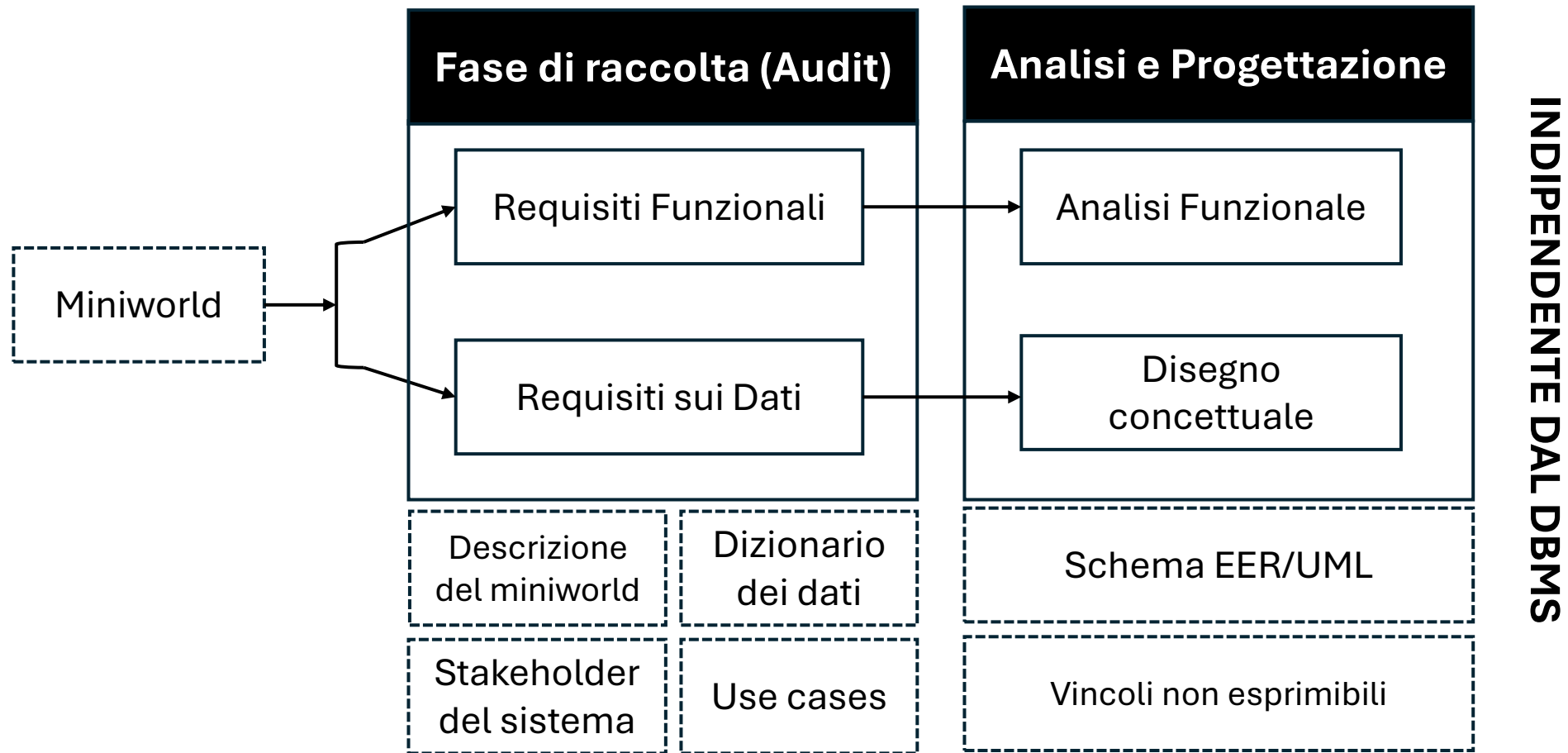
BASI DI DATI II

Progettare basi di dati di qualità: regole, best practice e linee guida

Anno 2024/2025
Prof. Genny Tortora
Dott. Luigi Di Biasi

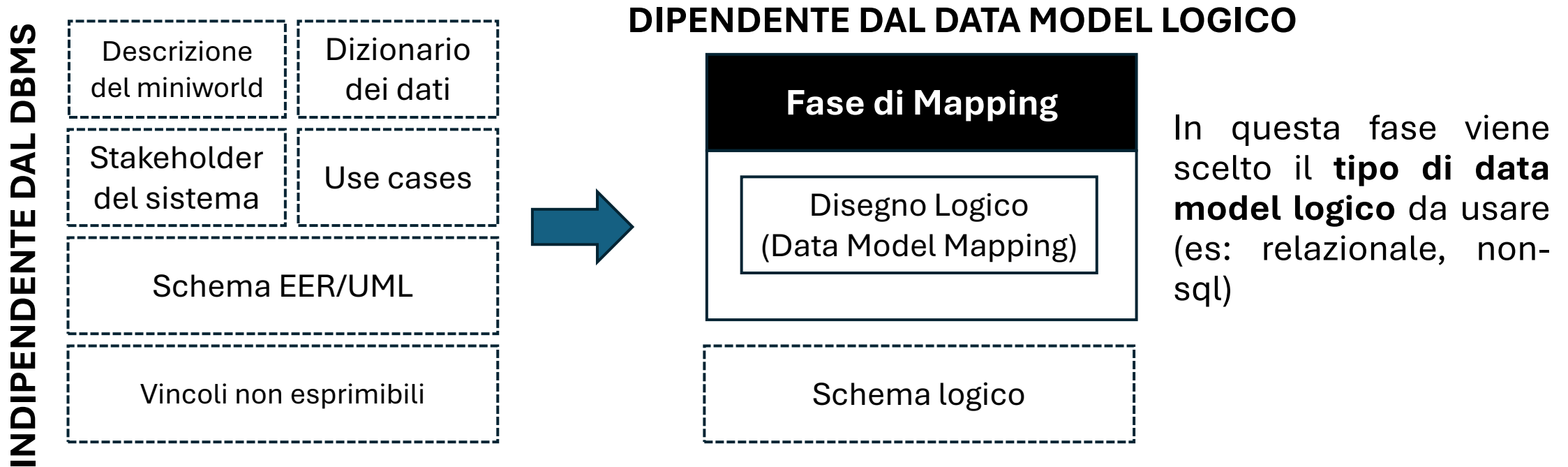
Progettazione di un database generico

Possiamo scomporre la progettazione di un database **tre fasi distinte**, una **indipendente** dal DBMS target, una indipendente dal DBMS target ma dipendente dal data-model scelto e una **dipendente** dal DBMS.



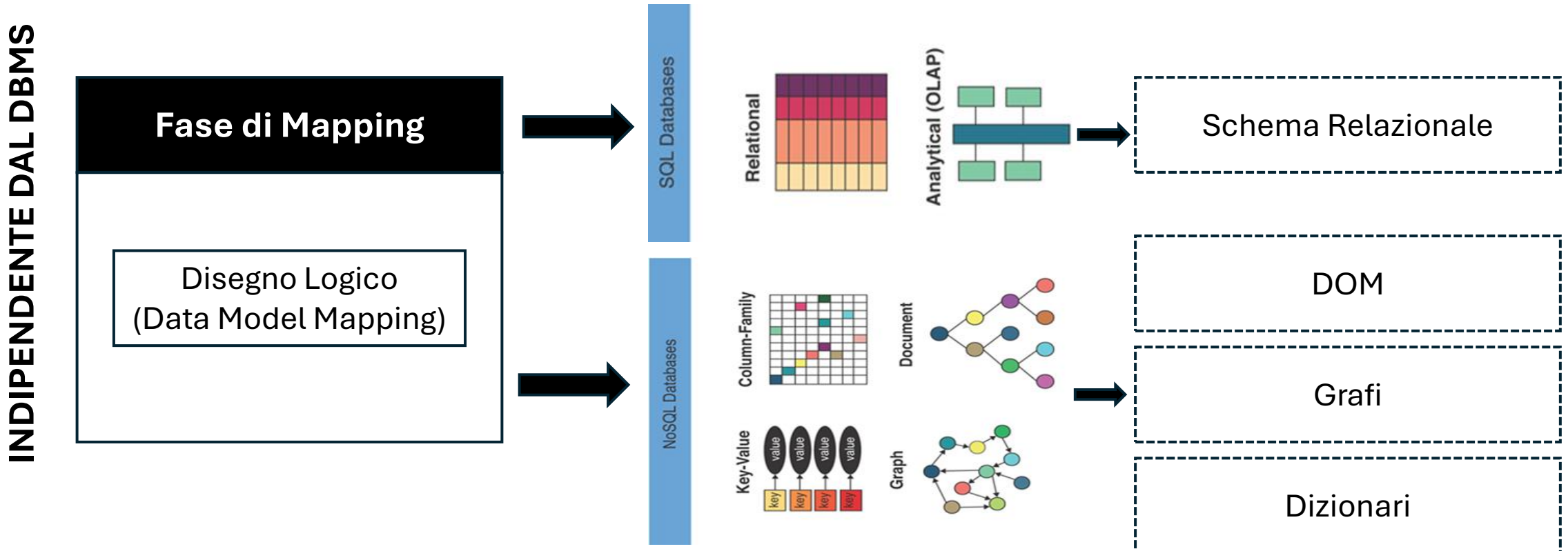
Progettazione di un database generico

Le informazioni estraibili dal miniworld vengono prima rappresentate mediante **un datamodel di alto livello (UML, EER, ...)**. Ciò permette di poter riutilizzare la fase di analisi nel momento in cui volessimo cambiare DBMS target.



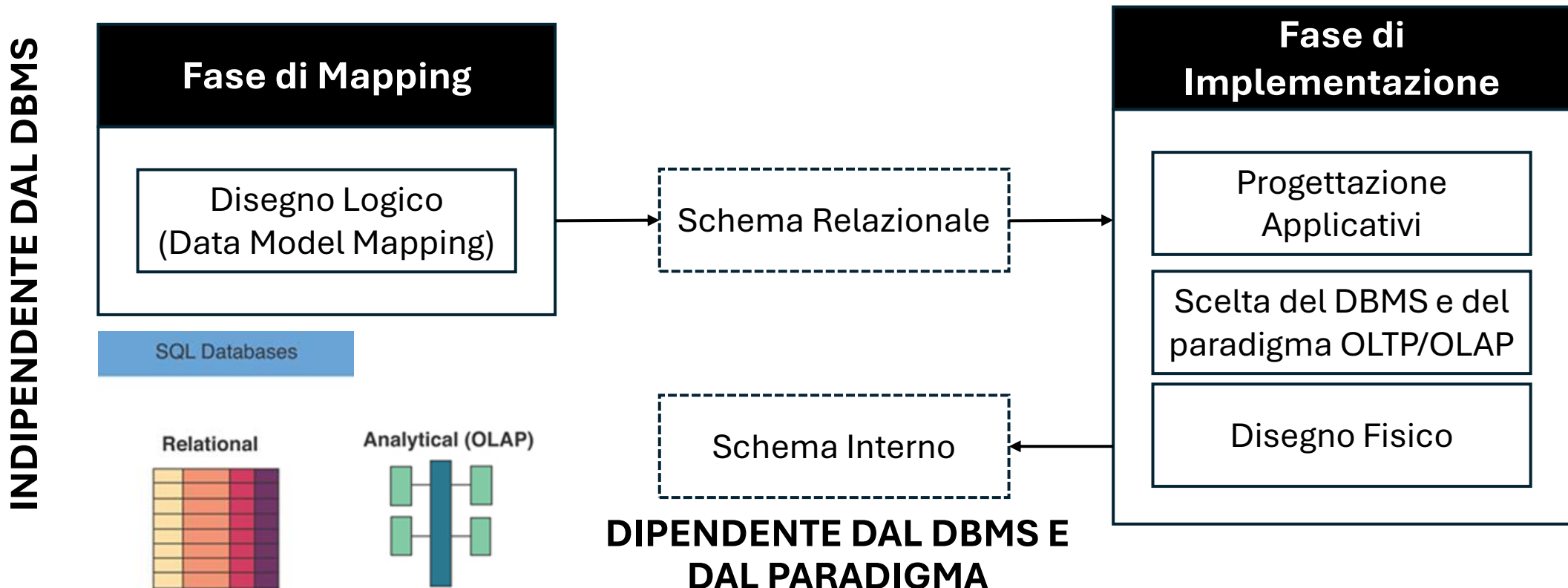
Progettazione di un database generico

Le informazioni estraibili dal miniworld vengono prima rappresentate mediante **un datamodel di alto livello (UML, EER, ...)**. Ciò permette di poter riutilizzare la fase di analisi nel momento in cui volessimo cambiare DBMS target.



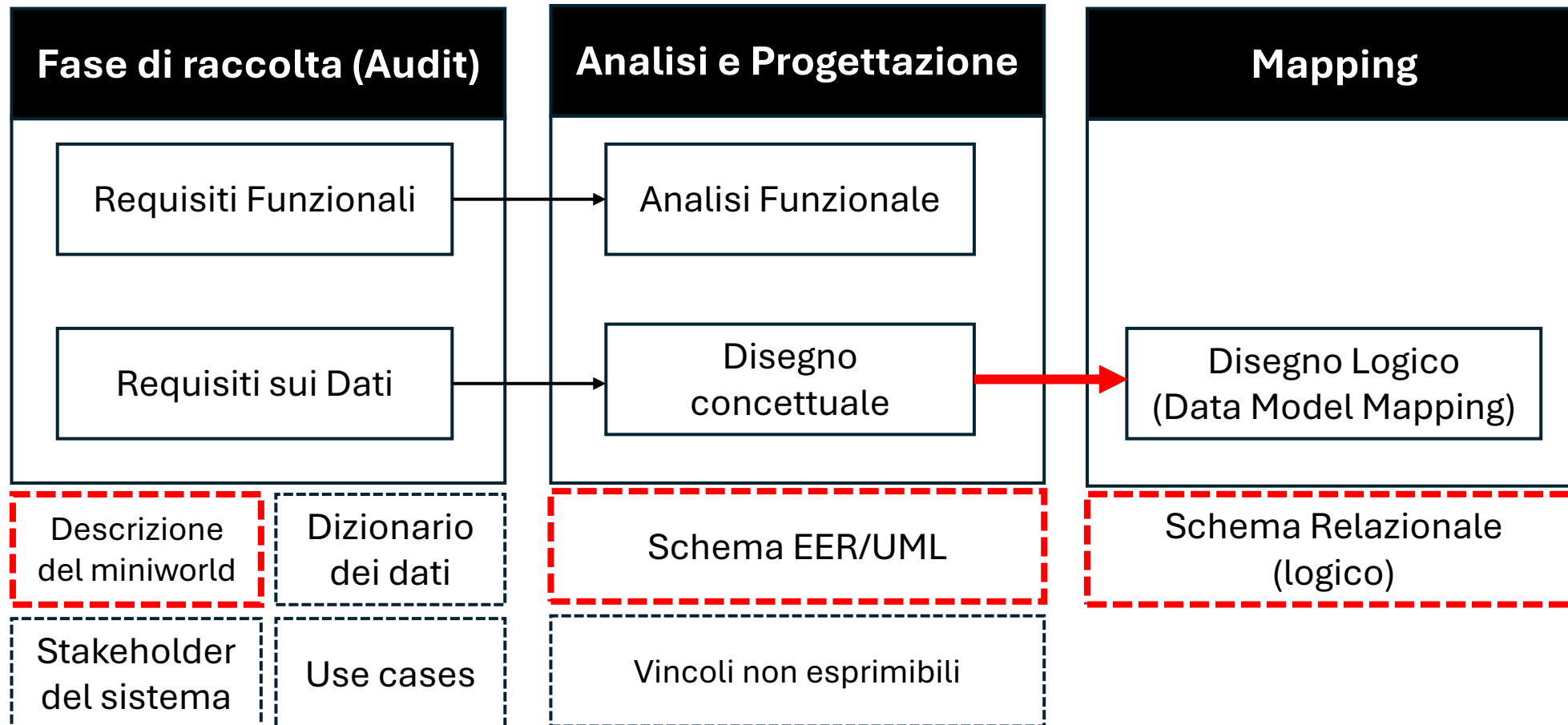
Progettazione di un database relazionale

Le informazioni estraibili dal miniworld vengono prima rappresentate mediante **un datamodel di alto livello (UML, EER, ...)**. Ciò permette di poter riutilizzare la fase di analisi nel momento in cui volessimo cambiare DBMS target.



Qualità degli schemi relazionali

La progettazione può presentare difetti, specialmente nella fase di mapping da ER a Relazionale. Questi errori possono generare **anomalie** nella base di dati.



Tipi di anomalie

Le “update anomalies” sono delle anomalie che possono sorgere nel gestire i dati di un database relazionale non correttamente progettato.

- **Insertion anomalies** ovvero, difficoltà nell'inserire nuovi dati senza informazioni ridondanti o incomplete;
- **Modification anomalies**; in questo caso, la modifica di un dato richiede cambiamenti multipli, aumentando il rischio di inconsistenze;
- **Deletion anomalies**– L'eliminazione di un dato può comportare la perdita di informazioni utili non duplicabili altrove.

Queste anomalie sono spesso causate da **ridondanze o una cattiva normalizzazione** dello schema relazionale.

Ricordate da Basi di dati 1, che tipologie di operazioni applicavamo allo schema EER prima di passare all'implementazione vera e propria?

Esempio di Insertion e modification anomaly

Qualcuno (*da licenziare*) ha progettato una base usando un unico schema di relazione EMP_DEPT per rappresentare sia le informazioni degli impiegati e sia delle (*TUTTE, TROPPE*) informazioni relative al dipartimento in cui lavora.

EMP_DEPT

ENAME	<u>SSN</u>	BDATE	ADDRESS	DNUMBER	DNAME	DMGRSSN
-------	------------	-------	---------	---------	-------	---------

Dovrebbe apparire chiaro che uno schema di relazione fatto in questo modo **ci espone a rischi di inconsistenza elevati**.

- Supponiamo di inserire Luigi e Antonio nel dipartimento 0, nel momento in cui Federico è il responsabile del dipartimento. Le tuple inserite assumerebbero la forma
 - (Luigi,DBSLGU,8885,Perdifumo,0,Laboratorio,**Federico**)
 - (Antonio,BNCANT,1234,Salerno,0,Laboratorio,**Federico**)
- Supponiamo che il responsabile cambi in Andrea e che un nuovo impiegato debba essere inserito: (Mara,XXXMAR,0123,Napoli,0,Laboratorio,**Andrea**)

Potremmo avere più impiegati nello stesso dipartimento con responsabili diversi. Inoltre, per gli impiegati senza dipartimento, dovremmo usare valori NULL. Se cambiasse un attributo del dipartimento, sarebbe necessario aggiornare tutte le tuple degli impiegati associati.

Esempio di deletion anomaly

Consideriamo ancora lo schema di relazione precedente. Se cancelliamo da EMP_DEPT una tupla impiegato che rappresenta l'ultimo impiegato che lavora per un particolare dipartimento, l'informazione sul dipartimento è persa dal database.

EMP_DEPT						
ENAME	<u>SSN</u>	BDATE	ADDRESS	DNUMBER	DNAME	DMGRSSN

Misure di bontà (Il vecchio testamento)

Quando si valuta **la qualità di uno schema di relazione (R) in un database**, bisogna considerare due aspetti fondamentali: la bontà a **livello logico** e la bontà a **livello di memorizzazione e manipolazione**.

Il livello logico **dello schema R riguarda il significato dei dati e il modo in cui sono organizzati**. Lo schema R deve essere chiaro e intuitivo, in modo che gli utenti possano capire facilmente quali informazioni sono contenute e come devono essere interrogate (query). Se uno schema è poco chiaro, gli utenti potrebbero fraintendere i dati o scrivere query errate.

Il livello di memorizzazione e manipolazione **si occupa di come i dati vengono effettivamente archiviati e aggiornati nel database**. Lo schema R **deve evitare problemi come la ridondanza** (ossia la ripetizione inutile di dati) e **garantire che le modifiche siano facili da applicare senza causare incoerenze**. Questo aspetto è particolarmente importante per le relazioni base, ovvero le tabelle fondamentali del database.

I 7 comandamenti del progettista relazionale

Misure di bontà informali

1. Fai in modo di avere una **semantica chiara degli attributi**;
2. **Riduci** il numero di valori **ridondanti** nelle tuple;
3. **Riduci** il numero di valori **null** nelle tuple;
4. **Non consentire tuple spurie**;

Linee guida

5. Disegna uno schema di relazione facile da spiegare;
6. Disegna gli schemi di relazione in modo che non possano accadere insertion, deletion o modification anomalies.
7. Progetta gli schemi in modo da poter eseguire JOIN senza generare tuple spurie.

Attenzione: Talvolta i comandamenti possono essere infranti per scopi di efficienza (lo vedremo quando parleremo di OLAP).

I° comandamento: «Fai in modo di avere una semantica chiara degli attributi»

EMPLOYEE

ENAME	<u>SSN</u>	BDATE	ADDRESS	DNUMBER
Smith,John B.	123456789	1965-01-09	731 Fondren,Houston,TX	5
Wong,Franklin T.	333445555	1955-12-08	638 Voss,Houston,TX	5
Zelaya,Alicia J.	999887777	1968-07-19	3321 Castle,Spring,TX	4
Wallace,Jennifer S.	987654321	1941-06-20	291 Berry,Bellaire,TX	4
Narayan,Remesh K.	666884444	1962-09-15	975 Fire Oak,Humble,TX	5
English,Joyce A.	453453453	1972-07-31	5631 Rice,Houston,TX	5
Jabbar,Ahmad V.	987987987	1969-03-29	980 Dallas,Houston,TX	4
Borg,James E.	888665555	1937-11-10	450 Stone,Houston,TX	1

DEPARTMENT

DNAME	<u>DNUMBER</u>	DMGRSSN
Research	5	333445555
Administration	4	987654321
Headquarters	1	888665555

DEPT_LOCATIONS

DNUMBER	DLOCATION
1	Houston
4	Stafford
5	Bellaire
5	Sugarland
5	Houston

WORKS_ON

<u>SSN</u>	<u>PNUMBER</u>	HOURS
123456789	1	32.5
123456789	2	7.5
666884444	3	40.0
453453453	1	20.0
453453453	2	20.0
333445555	2	10.0
333445555	3	10.0
333445555	10	10.0
333445555	20	10.0
999887777	30	30.0
999887777	10	10.0
987987987	10	35.0
987987987	30	5.0
987654321	30	20.0
987654321	20	15.0
888665555	20	null

PROJECT

PNAME	<u>PNUMBER</u>	PLOCATION	DNUM
ProductX	1	Bellaire	5
ProductY	2	Sugarland	5
ProductZ	3	Houston	5
Computerization	10	Stafford	4
Reorganization	20	Houston	1
Newbenefits	30	Stafford	4

Il significato, o **semantica**, **specifica come interpretare i valori** degli attributi di una relazione.

Più è semplice spiegare la semantica della relazione, **migliore è** il disegno dello schema di relazione

I° comandamento: «Fai in modo di avere una semantica chiara degli attributi»

EMPLOYEE

ENAME	<u>SSN</u>	BDATE	ADDRESS	DNUMBER
Smith,John B.	123456789	1965-01-09	731 Fondren,Houston,TX	5
Wong,Franklin T.	333445555	1955-12-08	638 Voss,Houston,TX	5
Zelaya,Alicia J.	999887777	1968-07-19	3321 Castle,Spring,TX	4
Wallace,Jennifer S.	987654321	1941-06-20	291 Berry,Bellaire,TX	4
Narayan,Remesh K.	666884444	1962-09-15	975 Fire Oak,Humble,TX	5
English,Joyce A.	453453453	1972-07-31	5631 Rice,Houston,TX	5
Jabbar,Ahmad V.	987987987	1969-03-29	980 Dallas,Houston,TX	4
Borg,James E.	888665555	1937-11-10	450 Stone,Houston,TX	1

DEPARTMENT

DNAME	<u>DNUMBER</u>	DMGRSSN
Research	5	333445555
Administration	4	987654321
Headquarters	1	888665555

WORKS_ON

<u>SSN</u>	<u>PNUMBER</u>	HOURS
123456789	1	32.5
123456789	2	7.5
666884444	3	40.0
453453453	1	20.0
453453453	2	20.0
333445555	2	10.0
333445555	3	10.0
333445555	10	10.0
333445555	20	10.0
999887777	30	30.0
999887777	10	10.0
987987987	10	35.0
987987987	30	5.0
987654321	30	20.0
987654321	20	15.0
888665555	20	null

DEPT_LOCATIONS

DNUMBER	DLOCATION
1	Houston
4	Stafford
5	Bellaire
5	Sugarland
5	Houston

PROJECT

PNAME	<u>PNUMBER</u>	PLOCATION	DNUM
ProductX	1	Bellaire	5
ProductY	2	Sugarland	5
ProductZ	3	Houston	5
Computerization	10	Stafford	4
Reorganization	20	Houston	1
Newbenefits	30	Stafford	4

Il significato, o **semantica**, **specifica come interpretare i valori** degli attributi di una relazione.

Più è semplice spiegare la semantica della relazione, **migliore è** il disegno dello schema di relazione

ENAME, SSN, BDATE, ADDRESS → dati dell'impiegato.

DNUMBER → dipartimento per cui lavora.

DNUMBER (foreign key – f.k.) → relazione implicita tra EMPLOYEE e DEPARTMENT.

DEPARTMENT → un'entità dipartimento.

PROJECT → una entità progetto.

f.k. DMGRSSN (di DEPARTMENT) → correla il dipartimento all'impiegato che è suo manager.

1° comandamento: «Fai in modo di avere una semantica chiara degli attributi»

EMPLOYEE

ENAME	<u>SSN</u>	BDATE	ADDRESS	DNUMBER
Smith,John B.	123456789	1965-01-09	731 Fondren,Houston,TX	5
Wong,Franklin T.	333445555	1955-12-08	638 Voss,Houston,TX	5
Zelaya,Alicia J.	999887777	1968-07-19	3321 Castle,Spring,TX	4
Wallace,Jennifer S.	987654321	1941-06-20	291 Berry,Bellaire,TX	4
Narayan,Remesh K.	666884444	1962-09-15	975 Fire Oak,Humble,TX	5
English,Joyce A.	453453453	1972-07-31	5631 Rice,Houston,TX	5
Jabbar,Ahmad V.	987987987	1969-03-29	980 Dallas,Houston,TX	4
Borg,James E.	888665555	1937-11-10	450 Stone,Houston,TX	1

DEPARTMENT

DNAME	<u>DNUMBER</u>	DMGRSSN
Research	5	333445555
Administration	4	987654321
Headquarters	1	888665555

DEPT_LOCATIONS

DNUMBER	DLOCATION
1	Houston
4	Stafford
5	Bellaire
5	Sugarland
5	Houston

WORKS_ON

<u>SSN</u>	<u>PNUMBER</u>	HOURS
123456789	1	32.5
123456789	2	7.5
666884444	3	40.0
453453453	1	20.0
453453453	2	20.0
333445555	2	10.0
333445555	3	10.0
333445555	10	10.0
333445555	20	10.0
999887777	30	30.0
999887777	10	10.0
987987987	10	35.0
987987987	30	5.0
987654321	30	20.0
987654321	20	15.0
888665555	20	null

PROJECT

PNAME	<u>PNUMBER</u>	PLOCATION	DNUM
ProductX	1	Bellaire	5
ProductY	2	Sugarland	5
ProductZ	3	Houston	5
Computerization	10	Stafford	4
Reorganization	20	Houston	1
Newbenefits	30	Stafford	4

Ogni tupla in DEPT_LOCATIONS contiene un numero di dipartimento (DNUMBER) e una delle locazioni del dipartimento (DLOCATIONS).

Ogni tupla in WORKS_ON contiene l'SSN di un impiegato, il numero di progetto PNUMBER di uno dei progetti su cui lavora l'impiegato, e il numero di ore settimanali HOURS.

DEPT_LOCATIONS → rappresenta un attributo multivalued di DEPARTMENT.

WORKS_ON → rappresenta una relazione di cardinalità M:N tra EMPLOYEE e PROJECT.

E' stato facile? Si? Ok, dunque tutte le relazioni possono essere considerate “buone” dal punto di vista di avere una semantica chiara. (1° comandamento soddisfatto!)

Il° comandamento: «Riduci il numero di valori ridondanti nelle tuple»

EMP_DEPT

ENAME	<u>SSN</u>	BDATE	ADDRESS	DNUMBER	DNAME	DMGRSSN
Smith,John B.	123456789	1965-01-09	731 Fondren,Houston,TX	5	Research	333445555
Wong,Franklin T.	333445555	1955-12-08	638 Voss,Houston,TX	5	Research	333445555
Zelaya, Alicia J.	999887777	1968-07-19	3321 Castle,Spring,TX	4	Administration	987654321
Wallace,Jennifer S.	987654321	1941-06-20	291 Berry,Bellaire,TX	4	Administration	987654321
Narayan,Ramesh K.	666884444	1962-09-15	975 FireOak,Humble,TX	5	Research	333445555
English,Joyce A.	453453453	1972-07-31	5631 Rice,Houston,TX	5	Research	333445555
Jabbar,Ahmad V.	987987987	1969-03-29	980 Dallas,Houston,TX	4	Administration	987654321
Borg,James E.	888665555	1937-11-10	450 Stone,Houston,TX	1	Headquarters	888665555

Oltre **allo spreco di spazio**, le relazioni EMP_DEPT e EMP_PROJ usate come relazioni base, presentano anche il problema delle **update anomalies**.

EMP_PROJ

<u>SSN</u>	<u>PNUMBER</u>	HOURS	ENAME	PNAME	PLOCATION
123456789	1	32.5	Smith,John B.	ProductX	Bellaire
123456789	2	7.5	Smith,John B.	ProductY	Sugarland
666884444	3	40.0	Narayan,Ramesh K.	ProductZ	Houston
453453453	1	20.0	English,Joyce A.	ProductX	Bellaire
453453453	2	20.0	English,Joyce A.	ProductY	Sugarland
333445555	2	10.0	Wong,Franklin T.	ProductY	Sugarland
333445555	3	10.0	Wong,Franklin T.	ProductZ	Houston
333445555	10	10.0	Wong,Franklin T.	Computerization	Stafford
333445555	20	10.0	Wong,Franklin T.	Reorganization	Houston
999887777	30	30.0	Zelaya,Alicia J.	Newbenefits	Stafford
999887777	10	10.0	Zelaya,Alicia J.	Computerization	Stafford
987987987	10	35.0	Jabbar,Ahmad V.	Computerization	Stafford
987987987	30	5.0	Jabbar,Ahmad V.	Newbenefits	Stafford
987654321	30	20.0	Wallace,Jennifer S.	Newbenefits	Stafford
987654321	20	15.0	Wallace,Jennifer S.	Reorganization	Houston
888665555	20	null	Borg,James E.	Reorganization	Houston

Come possiamo modificare questo schema di relazione per soddisfare il 2° comandamento?

Il° comandamento: «Riduci il numero di valori ridondanti nelle tuple»

EMP_DEPT						
ENAME	SSN	BDATE	ADDRESS	DNUMBER	DNAME	DMGRSSN
Smith, John B.	123456789	1965-01-09	731 Fondren, Houston, TX	5	Research	333445555
Wong, Franklin T.	333445555	1955-12-08	638 Voss, Houston, TX	5	Research	333445555
Zelaya, Alicia J.	999887777	1968-07-19	3321 Castle, Spring, TX	4	Administration	987654321
Wallace, Jennifer S.	987654321	1941-06-20	291 Berry, Bellaire, TX	4	Administration	987654321
Narayan, Ramesh K.	666884444	1962-09-15	975 Fire Oak, Humble, TX	5	Research	333445555
English, Joyce A.	453453453	1972-07-31	5631 Rice, Houston, TX	5	Research	333445555
Jabbar, Ahmad V.	987987987	1969-03-29	980 Dallas, Houston, TX	4	Administration	987654321
Borg, James E.	888665555	1937-11-10	450 Stone, Houston, TX	1	Headquarters	888665555

EMP_PROJ					
SSN	PNUMBER	HOURS	ENAME	PNAME	PLOCATION
123456789	1	32.5	Smith, John B.	ProductX	Bellaire
123456789	2	7.5	Smith, John B.	ProductY	Sugarland
666884444	3	40.0	Narayan, Ramesh K.	ProductZ	Houston
453453453	1	20.0	English, Joyce A.	ProductX	Bellaire
453453453	2	20.0	English, Joyce A.	ProductY	Sugarland
333445555	2	10.0	Wong, Franklin T.	ProductY	Sugarland
333445555	3	10.0	Wong, Franklin T.	ProductZ	Houston
333445555	10	10.0	Wong, Franklin T.	Computerization	Stafford
333445555	20	10.0	Wong, Franklin T.	Reorganization	Houston
999887777	30	30.0	Zelaya, Alicia J.	Newbenefits	Stafford
999887777	10	10.0	Zelaya, Alicia J.	Computerization	Stafford
987987987	10	35.0	Jabbar, Ahmad V.	Computerization	Stafford
987987987	30	5.0	Jabbar, Ahmad V.	Newbenefits	Stafford
987654321	30	20.0	Wallace, Jennifer S.	Newbenefits	Stafford
987654321	20	15.0	Wallace, Jennifer S.	Reorganization	Houston
888665555	20	null	Borg, James E.	Reorganization	Houston



EMPLOYEE				
ENAME	SSN	BDATE	ADDRESS	DNUMBER
Smith, John B.	123456789	1965-01-09	731 Fondren, Houston, TX	5
Wong, Franklin T.	333445555	1955-12-08	638 Voss, Houston, TX	5
Zelaya, Alicia J.	999887777	1968-07-19	3321 Castle, Spring, TX	4
Wallace, Jennifer S.	987654321	1941-06-20	291 Berry, Bellaire, TX	4
Narayan, Ramesh K.	666884444	1962-09-15	975 Fire Oak, Humble, TX	5
English, Joyce A.	453453453	1972-07-31	5631 Rice, Houston, TX	5
Jabbar, Ahmad V.	987987987	1969-03-29	980 Dallas, Houston, TX	4
Borg, James E.	888665555	1937-11-10	450 Stone, Houston, TX	1

DEPARTMENT		
DNAME	DNUMBER	DMGRSSN
Research	5	333445555
Administration	4	987654321
Headquarters	1	888665555

WORKS_ON		
SSN	PNUMBER	HOURS
123456789	1	32.5
123456789	2	7.5
666884444	3	40.0
453453453	1	20.0
453453453	2	20.0
333445555	2	10.0
333445555	3	10.0
333445555	10	10.0
333445555	20	10.0
999887777	30	30.0
999887777	10	10.0
987987987	10	35.0
987987987	30	5.0
987654321	30	20.0
987654321	20	15.0
888665555	20	null

DEPT_LOCATIONS	
DNUMBER	DLOCATION
1	Houston
4	Stafford
5	Bellaire
5	Sugarland
5	Houston

PROJECT			
PNAME	PNUMBER	PLOCATION	DNUM
ProductX	1	Bellaire	5
ProductY	2	Sugarland	5
ProductZ	3	Houston	5
Computerization	10	Stafford	4
Reorganization	20	Houston	1
Newbenefits	30	Stafford	4

Vi ricorda qualcosa questa operazione?

Suggerimento...

DNUMBER -> DNAME, DMGRSSN

SSN -> BDATE, ADDRESS, ENAME

SSN, PNUMBER -> HOURS, PNAME, PLOC

III° comandamento: «**Riduci** il numero di valori **null** nelle tuple»

Il valore **null** può avere diverse interpretazioni: L'attributo non si applica a questa tupla; Il valore dell'attributo per questa tupla non è noto (i famosi **NA**); Il valore dell'attributo è noto ma assente, cioè non è stato ancora registrato.

A volte capita, in alcuni disegni di schemi, **di raggruppare molti attributi in una singola relazione grossa**. Purtroppo, se molti degli attributi non si applicano a tutte le tuple possiamo avere molti valori **null**. Oltre allo spreco di spazio, ciò crea problemi con le funzioni di aggregazione COUNT e SUM.

Quello che possiamo fare è «Evitare (*per quanto possibile*) di porre attributi in una relazione base i cui valori possono essere null.»

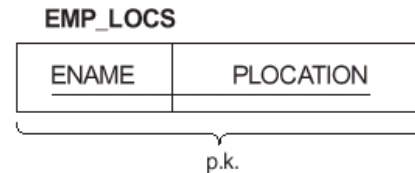
Se i valori **null** sono inevitabili, **dobbiamo assicurarci** che essi ricorrano in casi eccezionali e non per la maggioranza delle tuple. Ad esempio, se solo il 10% degli impiegati ha un ufficio individuale, è poco giustificato include l'attributo NUMERO_UFFICIO nella relazione IMPIEGATI.

Ha più senso creare la relazione:

IMPIEGATI_UFFICI (ESSN, OFFICE_NUMBER)

IV° comandamento: «Non consentire tuple spurie»

Per focalizzare bene il problema, consideriamo gli schemi di relazione seguenti.



L'impiegato di nome ENAME lavora per qualche progetto la cui locazione è PLOCATION.



EMP_PROJ1: l'impiegato con codice SSN lavora HOURS ore alla settimana sul progetto avente numero PNUMBER, nome PNAME e locazione PLOCATION.

La suddivisione non corrisponde ad un buon disegno di db in quanto, volendo riottenere le informazioni di EMP_PROJ mediante un natural join sui due schemi, si ottengono **tuple spurie**, rappresentanti informazioni errate.

IV° comandamento: «Non consentire tuple spurie»

Provate a fare una natural join tra queste due tabelle.

EMP_LOCS		EMP_PROJ1				
ENAME	PLOCATION	SSN	PNUMBER	HOURS	PNAME	PLOCATION
Smith, John B.	Bellaire	123456789	1	32.5	Product X	Bellaire
Smith, John B.	Sugarland	123456789	2	7.5	Product Y	Sugarland
Narayan, Ramesh K.	Houston	666884444	3	40.0	Product Z	Houston
English, Joyce A.	Bellaire	453453453	1	20.0	Product X	Bellaire
English, Joyce A.	Sugarland	453453453	2	20.0	Product Y	Sugarland
Wong, Franklin T.	Sugarland	333445555	2	10.0	Product Y	Sugarland
Wong, Franklin T.	Houston	333445555	3	10.0	Product Z	Houston
Wong, Franklin T.	Stafford	333445555	10	10.0	Computerization	Stafford
		333445555	20	10.0	Reorganization	Houston

Ricordiamo che **la natural join confronta automaticamente tutte le colonne con lo stesso nome nelle due tabelle** e restituisce solo le righe in cui i valori di queste colonne corrispondono.

In EMP_LOCS **Bellaire** è associato a Smith, John B e English, Joyce A, mentre in EMP_PROJ1 è associato a 123456789,453453453. Dopo la natural join cosa accade?

IV° comandamento: «Non consentire tuple spurie»

	SSN	PNUMBER	HOURS	PNAME	PLOCATION	
	123456789	1	32.5	ProductX	Bellaire	Smith,John B.
*	123456789	1	32.5	ProductX	Bellaire	English,Joyce A.
	123456789	2	7.5	ProductY	Sugarland	Smith,John B.
*	123456789	2	7.5	ProductY	Sugarland	English,Joyce A.
*	123456789	2	7.5	ProductY	Sugarland	Wong,Franklin T.
	666884444	3	40.0	ProductZ	Houston	Narayan,Ramesh K.
*	666884444	3	40.0	ProductZ	Houston	Wong,Franklin T.
*	453453453	1	20.0	ProductX	Bellaire	Smith,John B.
	453453453	1	20.0	ProductX	Bellaire	English,Joyce A.
*	453453453	2	20.0	ProductY	Sugarland	Smith,John B.
	453453453	2	20.0	ProductY	Sugarland	English,Joyce A.

Due persone diverse con lo stesso SSN!

Due righe doppiate con solo il valore di ENAME a fare da discriminante.

Tale problema sorge perché PLOCATION è l'attributo che correla EMP_LOCS e EMP_PROJ1, **ma non è né chiave primaria, né chiave esterna** in nessuna delle due relazioni.

V° comandamento: «Disegna uno schema di relazione facile da spiegare»

Non vanno mai combinati attributi da tipi di entità e tipi di relazioni **differenti** in una singola relazione.

Idealmente: ogni relazione dovrebbe essere «self-contained». Di conseguenza, **uno** schema di relazione **deve corrispondere esattamente ad un** tipo di entità o ad **un** tipo di relazione.

Esempio (negativo, da non seguire)

EMP_DEPT						
ENAME	<u>SSN</u>	BDATE	ADDRESS	DNUMBER	DNAME	DMGRSSN

EMP_PROJ					
<u>SSN</u>	<u>PNUMBER</u>	HOURS	ENAME	PNAME	PLOCATION

Questi due schemi di relazione hanno una semantica chiara, però entrambi contravvengono al V° comandamento, contenendo attributi di entità distinte.

VI° comandamento: «Disegna schemi che evitino insertion, deletion o modification anomalies»

Purtroppo, non sempre è possibile. Dunque non seguirlo lo consideriamo un «peccato veniale»

Se qualche anomalia è presente, è necessario annotare ciò chiaramente **in modo che i programmi (l'integrità e la coerenza dei dati viene demandata al programmatore - attenzione: è una deroga!)** che aggiornano il database operino correttamente.

Una soluzione migliore potrebbe comunque essere quella di definire delle viste.

VII° comandamento: «Progetta gli schemi per poter eseguire JOIN senza generare tuple spurie.»

REGOLA: Dobbiamo progettare gli schemi di relazione **in modo da poter effettuare JOIN con condizioni di uguaglianza su attributi che sono o chiave primaria o chiave esterna**, in modo da non generare tuple spurie.

TRUCCO: Dobbiamo progettare gli schemi in modo da poter usare sempre una INNER JOIN per ricostruire tutte le informazioni che ci servono.

PAUSA