

UNIVERSITÀ DEGLI STUDI "ROMA 3"

CORSO DI LAUREA IN INGEGNERIA INFORMATICA

GAIA: DATA TRANSLATION REUSE



PROF. RICCARDO
TORLONE

PAOLO PAPOTTI

ALESSIO DE
NICHILLO

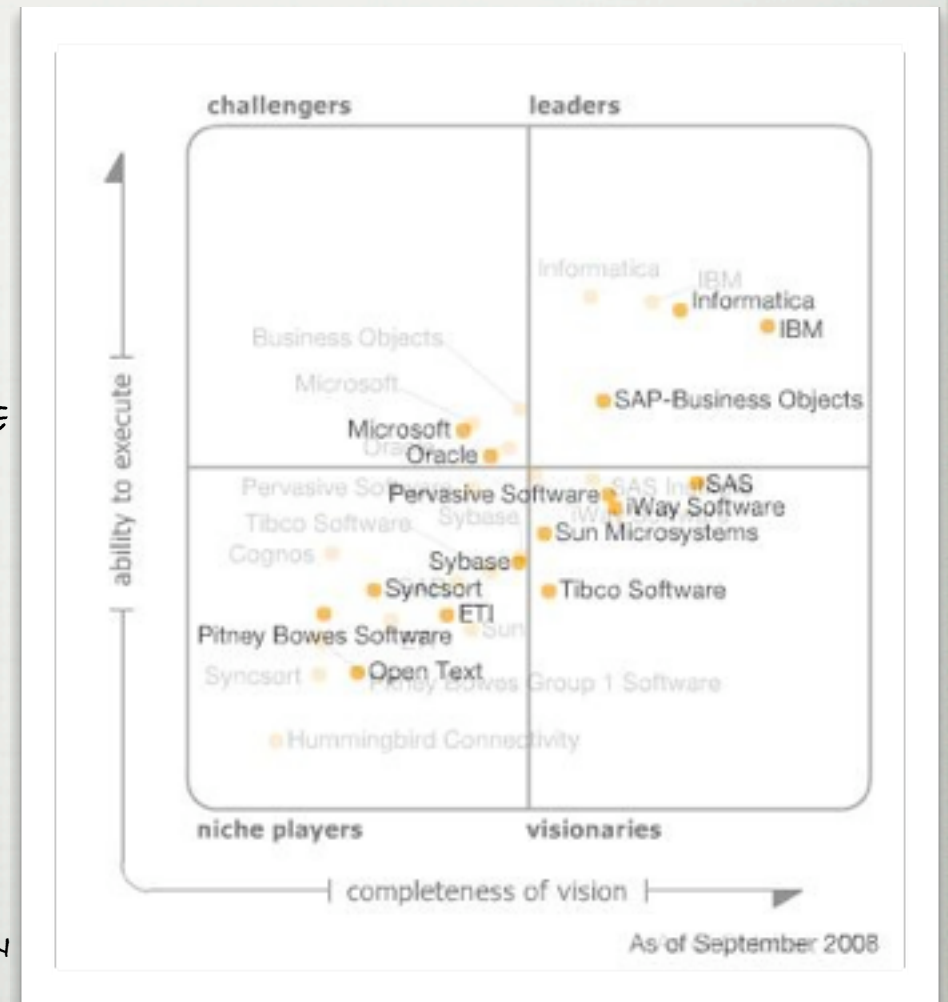
DATA INTEGRATION

- "68% OF THE CEOS LISTED THE INTEGRATION OF DISPARATE APPLICATIONS AND INFRASTRUCTURE AS A KEY ISSUE FOR THEIR BUSINESS, ONE THAT SLOWS THEM DOWN AND STOPS THE FLOW OF INFORMATION"

IBM BUSINESS CONSULTING SERVICES: YOUR TURN, THE GLOBAL CEO STUDY 2004.

- "79% OF COMPANIES (OF ALL SIZES) HAVE MORE THAN TWO DOCUMENT STORES, WHILE 25% HAVE MORE THAN FIFTEEN."

MOORE, C., MARKHAM, R.: THE FUTURE OF CONTENT IN THE ENTERPRISE. FORRESTER REPORT 2003.



DATA INTEGRATION

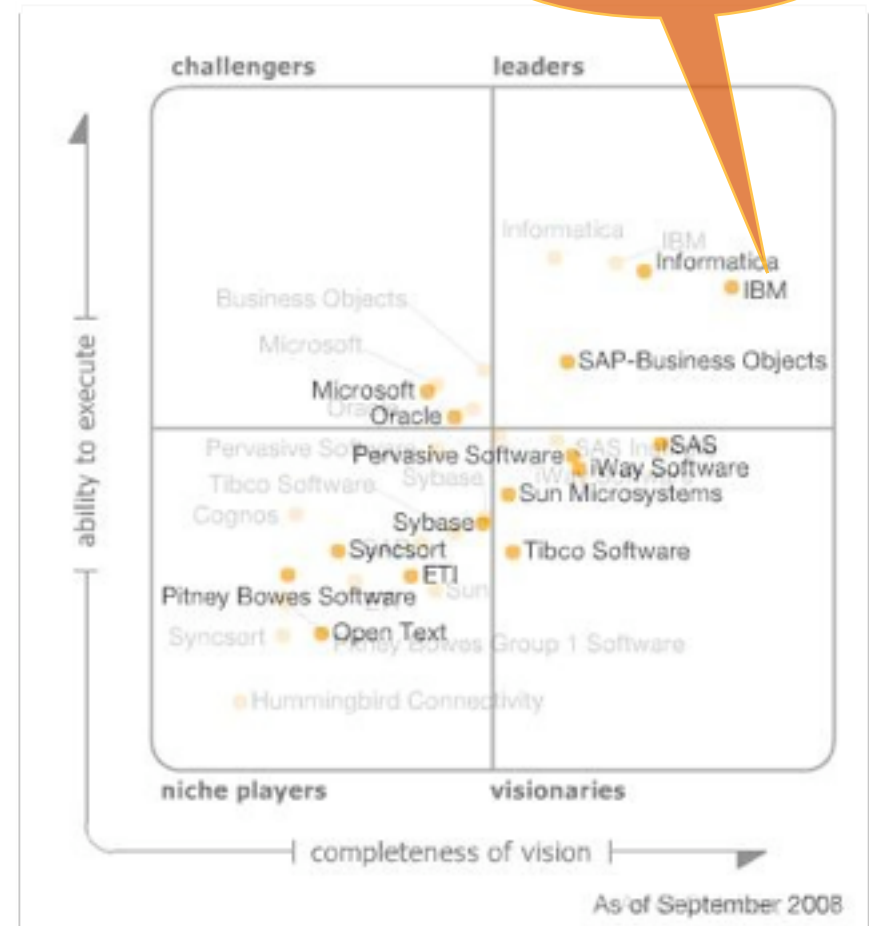
- "68% OF THE CEOS LISTED THE INTEGRATION OF DISPARATE APPLICATIONS AND INFRASTRUCTURE AS A KEY ISSUE FOR THEIR BUSINESS, ONE THAT SLOWS THEM DOWN AND STOPS THE FLOW OF INFORMATION"

IBM BUSINESS CONSULTING SERVICES: YOUR TURN, THE GLOBAL CEO STUDY 2004.

- "79% OF COMPANIES (OF ALL SIZES) HAVE MORE THAN TWO DOCUMENT STORES, WHILE 25% HAVE MORE THAN FIFTEEN."

MOORE, C., MARKHAM, R.: THE FUTURE OF CONTENT IN THE ENTERPRISE. FORRESTER REPORT 2003.

DATA EXCHANGE



CASO D'USO - I

MARIO ROSSI, DIPENDENTE DI UNA GRANDE COMPAGNIA, HA A DISPOSIZIONE LA SEGUENTE BASE DI DATI:

RICERCATORI

<u>nome</u>	titolo	gruppo
Davide	Dott.	IA
Agnese	Ing.	DB
Luca	Ing.	DB

GRUPPI

<u>id</u>	group
DB	database
IA	intellig. artific.

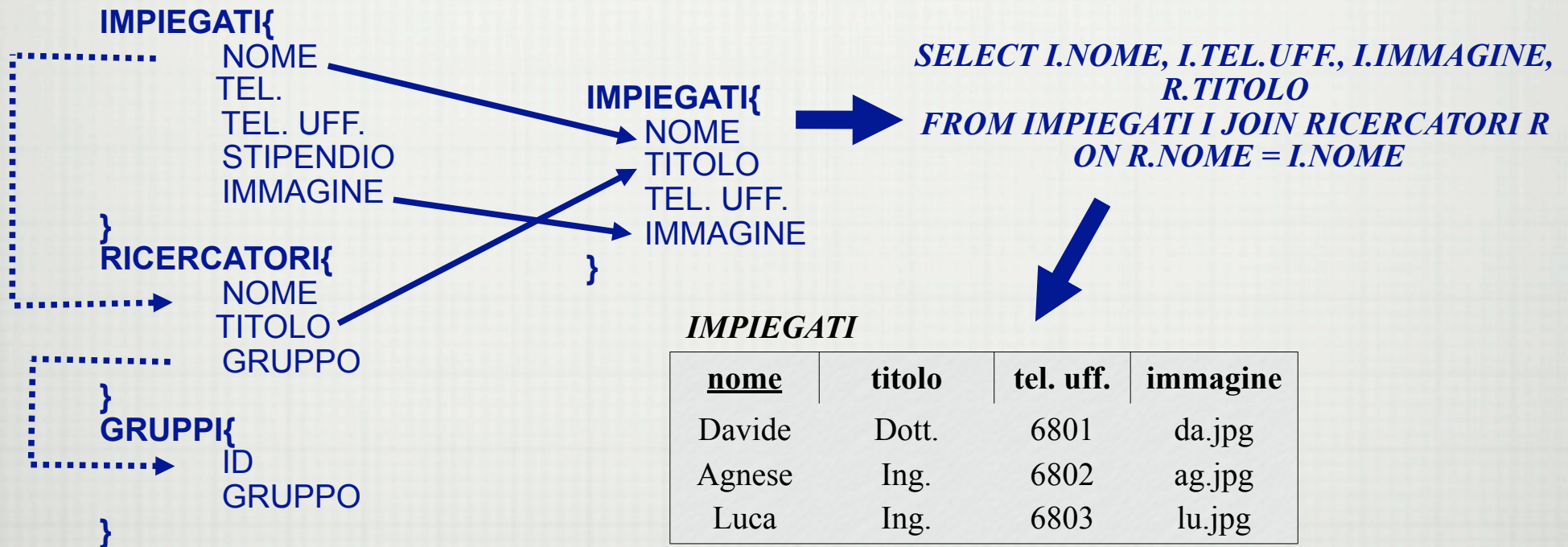
IMPIEGATI

<u>nome</u>	tel.	tel. uff.	stipendio	immagine
Davide	06636363	6801	55K	da.jpg
Agnese	06646464	6802	92K	ag.jpg
Luca	02545647	6803	67K	lu.jpg
Michele	06486313	6804	85K	mi.jpg
Angela	02678777	6805	67K	an.jpg
Bruno	07361233	6806	105K	br.jpg

SU RICHIESTA DI UN CLIENTE, DEVE FORNIRE UNA BASE DI DATI CHE CONTENGA ALCUNE INFORMAZIONI RIGUARDANTI I RICERCATORI CHE FANNO PARTE DELL'AZIENDA AD ESEMPIO NOME, TITOLO, NUMERO TELEFONICO DELL'UFFICIO ED IMMAGINE

CASO D'USO - 2

DEFINISCE UNA TRASFORMAZIONE CHE GLI PERMETTE DI RENDERE DISPONIBILI I DATI RICHIESTI NASCONDENDO PERÒ INFORMAZIONI SENSIBILI NON NECESSARIE



ESISTONO TOOL CHE PERMETTONO DI DEFINIRE GRAFICAMENTE UNA TRASFORMAZIONE DEL GENERE E GENERANO GLI SCRIPT SQL (XQUERY) ADATTI

CASO D'USO - 3

UN ALTRO DIPENDENTE DELLA SUDETTA COMPAGNIA, SI TROVA AD AFFRONTARE UN PROBLEMA SIMILE CON LA SEGUENTE BASE DI DATI, DIFFERENTE DALLA PRECEDENTE, MA SIMILE NELLA STRUTTURA:

SEDI

<u>nome</u>	città	progetto
Luca	Milano	prj1
Michele	Roma	prj1
Bruno	Ascoli	prj2

STABILIMENTO

<u>id</u>	nome
prj1	progetto uno
prj2	progetto2

IMPIEGATI

<u>nome</u>	tel.	tel. uff.	stipendio	immagine
Davide	06636363	6801	55K	da.jpg
Agnese	06646464	6802	92K	ag.jpg
Luca	02545647	6803	67K	lu.jpg
Michele	06486313	6804	85K	mi.jpg
Angela	02678777	6805	67K	an.jpg
Bruno	07361233	6806	105K	br.jpg

NON È POSSIBILE RIUTILIZZARE -AUTOMATICAMENTE- LO SCRIPT PRODOTTO IN PRECEDENZA!

DEVE RIPROGETTARE TUTTA LA TRASFORMAZIONE DA CAPO PER RIOTTENERE UN RISULTATO SIMILE?

CASO D'USO - 3

UN ALTRO DIPENDENTE DELLA SUDDETTA COMPAGNIA, SI TROVA AD AFFRONTARE UN PROBLEMA SIMILE CON LA SEGUENTE BASE DI DATI, DIFFERENTE DALLA PRECEDENTE, MA SIMILE NELLA STRUTTURA:

SEDI

<u>nome</u>	città	progetto
Luca	Milano	prj1
Michele	Roma	prj1
Bruno	Ascoli	prj2

STABILIMENTO

<u>id</u>	nome
prj1	progetto uno
prj2	progetto2

NON È POSSIBILE RIUTILIZZARE IL
PRODOTTO IN PRECEDENZA!
DEVE RIPROGETTARE TUTTA LA
STRUTTURA PER OTTENERE UN RISULTATO.

IMPIEGATI

<u>nome</u>	tel.	tel. uff.	stipendio	immagine
-------------	------	-----------	-----------	----------

RICERCATORI

<u>nome</u>	titolo	gruppo
<u>Davide</u>	Dott.	IA
<u>Agnese</u>	Ing.	DB
<u>Luca</u>	Ing.	DB

GRUPPI

<u>id</u>	group
DB	database
IA	intellig. artific.

IMPIEGATI

<u>nome</u>	tel.
<u>Davide</u>	0663636
<u>Agnese</u>	0664646
<u>Luca</u>	0254564
Michele	0648631
Angela	0267877
Bruno	0736123

CASO D'USO - 3

UN ALTRO DIPENDENTE DELLA SUDETTA COMPAGNIA, SI TROVA AD AFFRONTARE UN PROBLEMA SIMILE CON LA SEGUENTE BASE DI DATI, DIFFERENTE DALLA PRECEDENTE, MA SIMILE NELLA STRUTTURA:

SEDI

<u>nome</u>	città	progetto
Luca	Milano	prj1
Michele	Roma	prj1
Bruno	Ascoli	prj2

STABILIMENTO

<u>id</u>	nome
prj1	progetto uno
prj2	progetto2

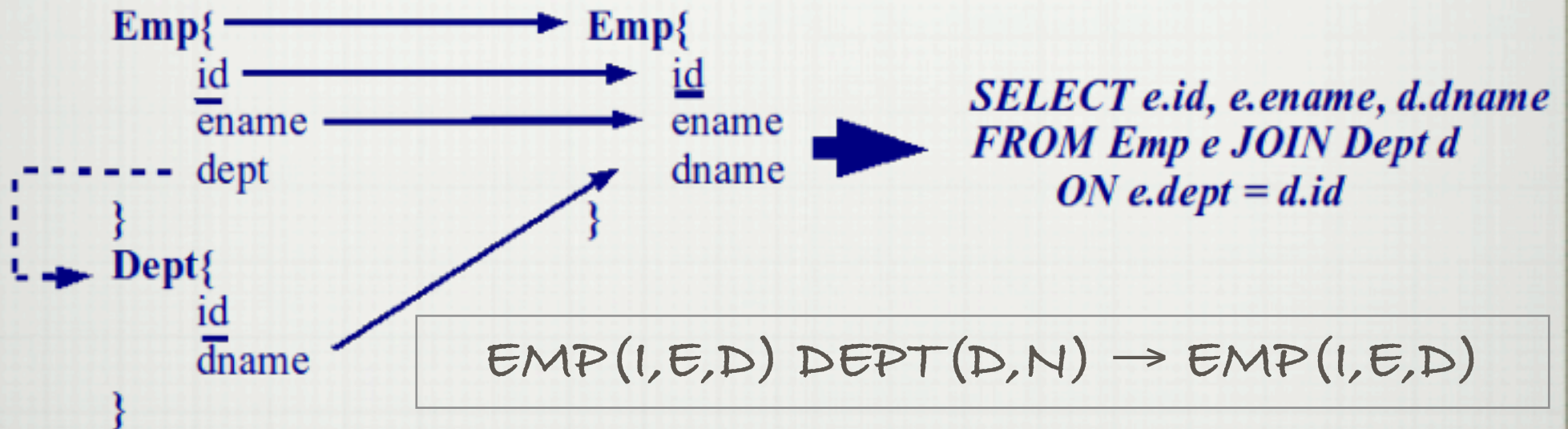
IMPIEGATI

<u>nome</u>	tel.	tel. uff.	stipendio	immagine
Davide	06636363	6801	55K	da.jpg
Agnese	06646464	6802	92K	ag.jpg
Luca	02545647	6803	67K	lu.jpg
Michele	06486313	6804	85K	mi.jpg
Angela	02678777	6805	67K	an.jpg
Bruno	07361233	6806	105K	br.jpg

NON È POSSIBILE RIUTILIZZARE -AUTOMATICAMENTE- LO SCRIPT PRODOTTO IN PRECEDENZA!
DEVE RIPROGETTARE TUTTA LA TRASFORMAZIONE DA CAPO PER RIOTTENERE UN RISULTATO SIMILE?

SOLUZIONI ESISTENTI

DATA EXCHANGE

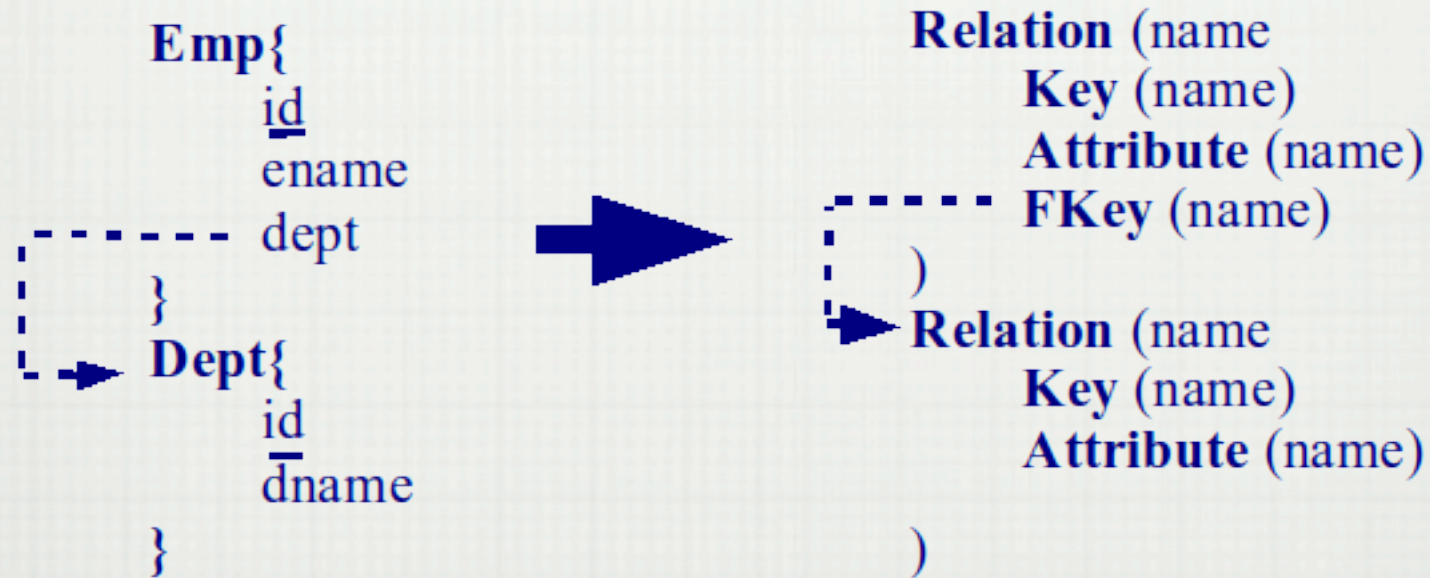


PROBLEMI

- GLI SCRIPT GENERATI SPESSO NON SONO RIUTILIZZABILI
- NON È BEN CHIARO IN CHE MODO ARCHIVIARLI E COME CERCARLI IN CASO DI RIUSO
- PERDITA DI ENTERPRISE KNOWLEDGE

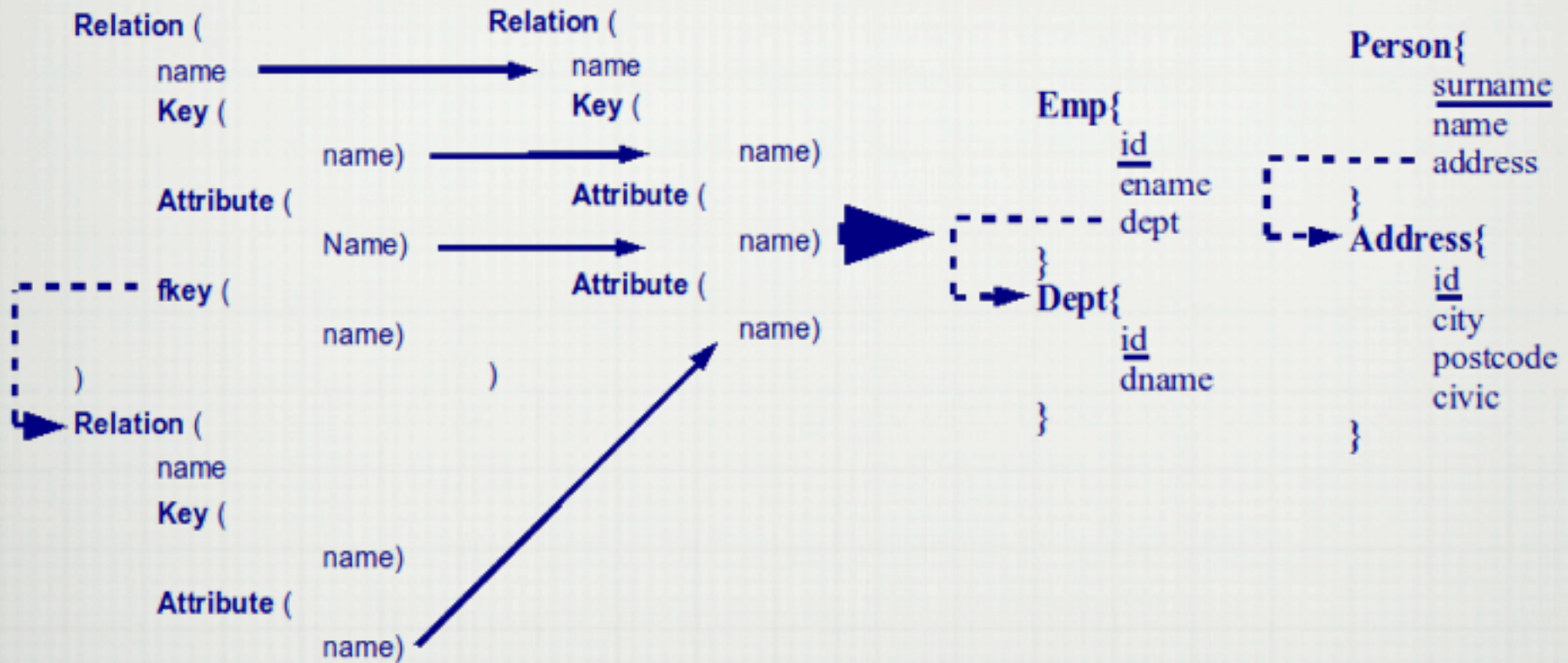
INNOVAZIONE

TEMPLATE DI SCHEMI DI BASI DI DATI



SI PASSA DA UNA DESCRIZIONE PARTICOLAREGGIATA DELLO SCHEMA AD UNA PIÙ GENERALE, CHE NE CATTURA ESATTAMENTE LA STRUTTURA, MA CHE SI PUÒ APPLICARE ANCHE A SCHEMI DIFFERENTI

INNOVAZIONE

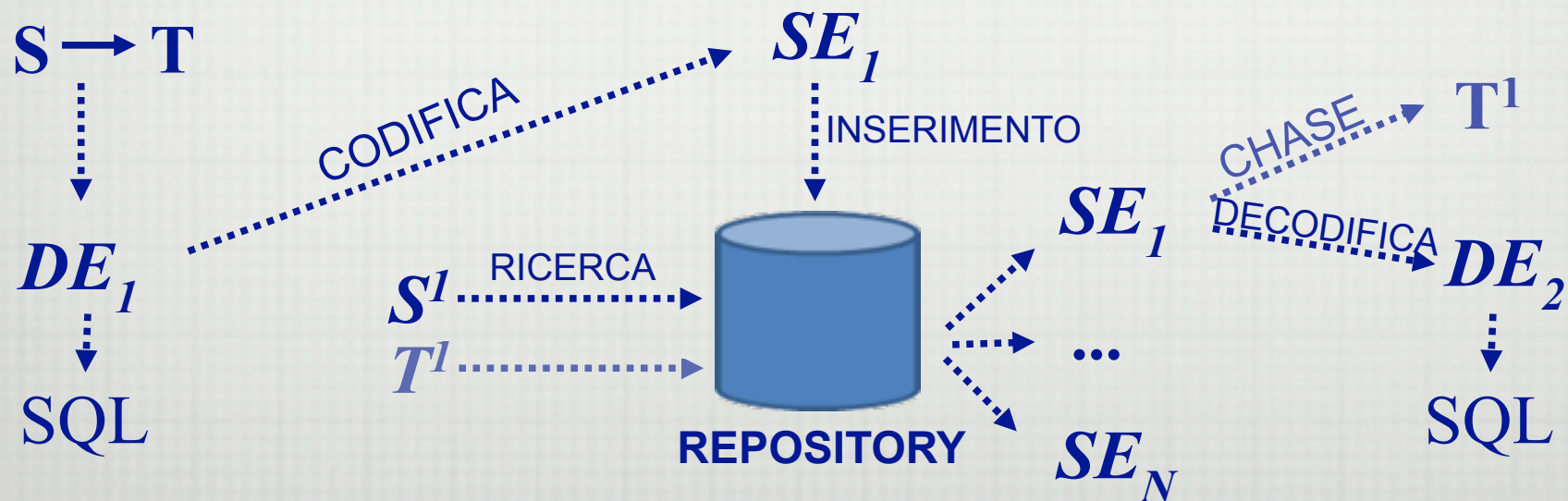


REL(R) KEY(K,R) ATT(A,R) FK(F,R,R') REL(R') KEY(K',R') ATT(A',R') →
 REL(R) KEY(K,R) ATT(A,R) ATT(A',R)

SCHEMA EXCHANGE: APPLICABILE A SCHEMI DIFFERENTI

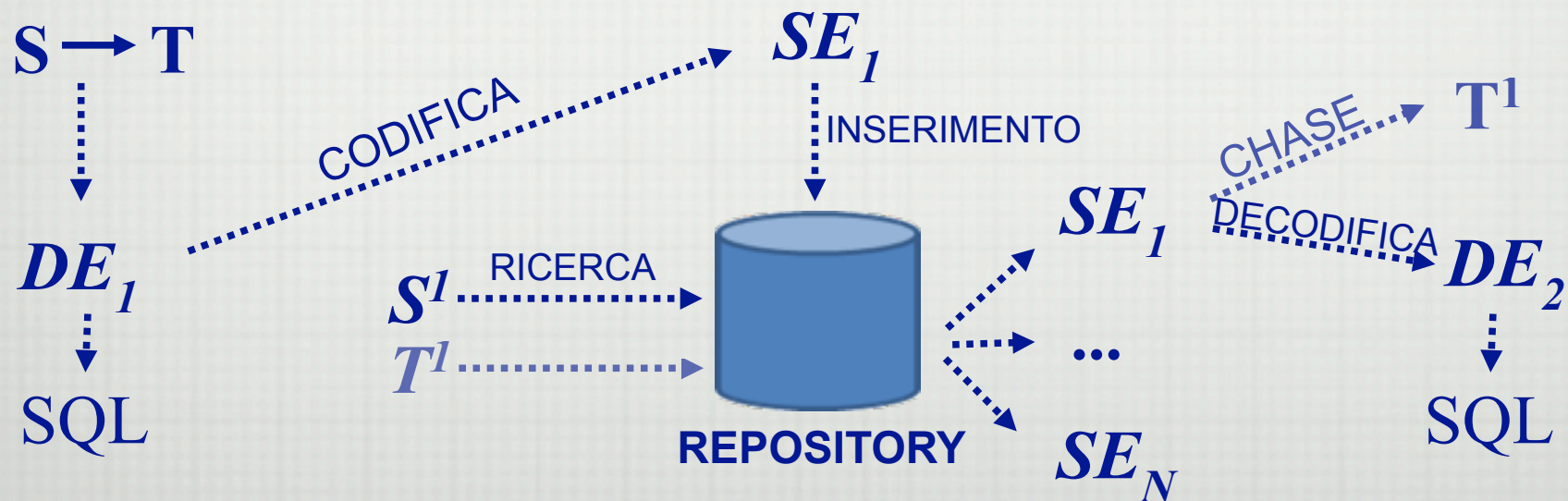
PROCESSO

1. DATO SCHEMA SOURCE S E SCHEMA TARGET T.



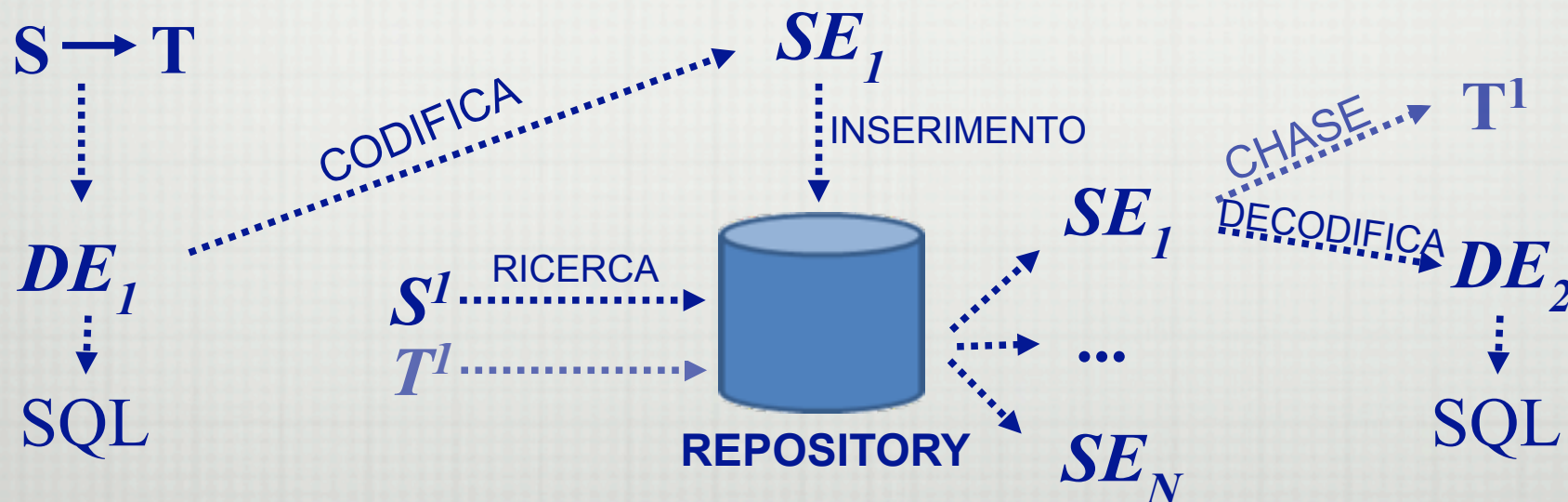
PROCESSO

1. DATO SCHEMA SOURCE S E SCHEMA TARGET T .
2. GENERA UN DATA EXCHANGE DE_1 CHE PRODUCE UNA TRASFORMAZIONE SUI DATI IN SQL



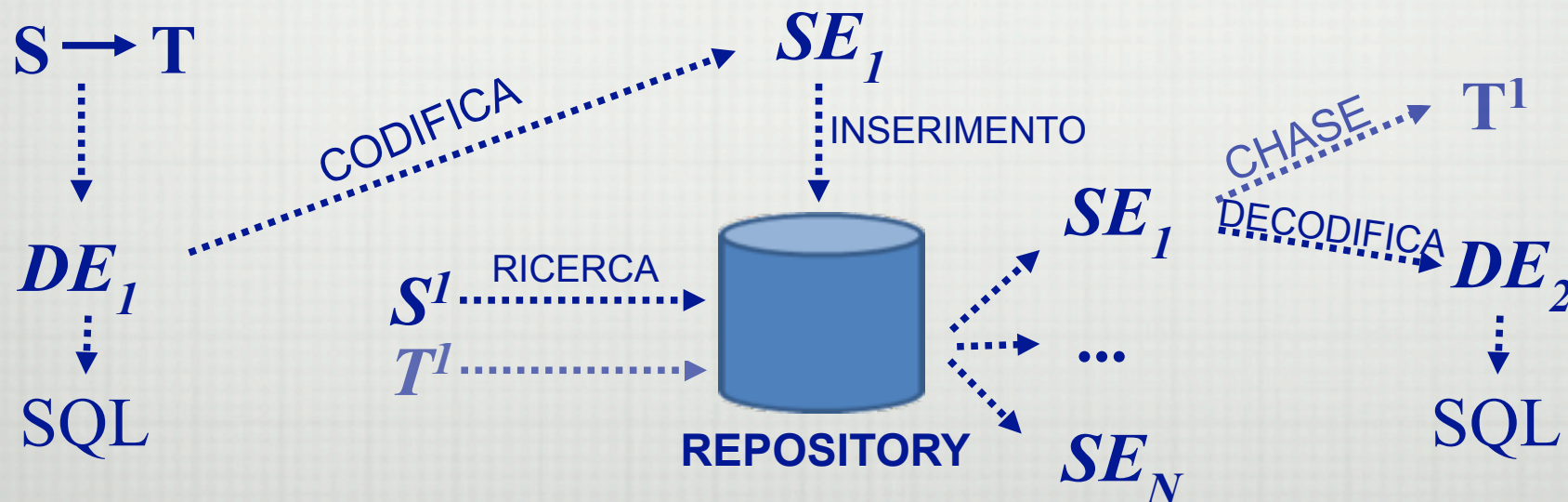
PROCESSO

1. DATO SCHEMA SOURCE S E SCHEMA TARGET T .
2. GENERA UN DATA EXCHANGE DE_1 CHE PRODUCE UNA TRASFORMAZIONE SUI DATI IN SQL
3. IL DATA EXCHANGE VIENE CODIFICATO IN UNO SCHEMA EXCHANGE SE_1 CHE VIENE INSERITO NEL REPOSITORY.



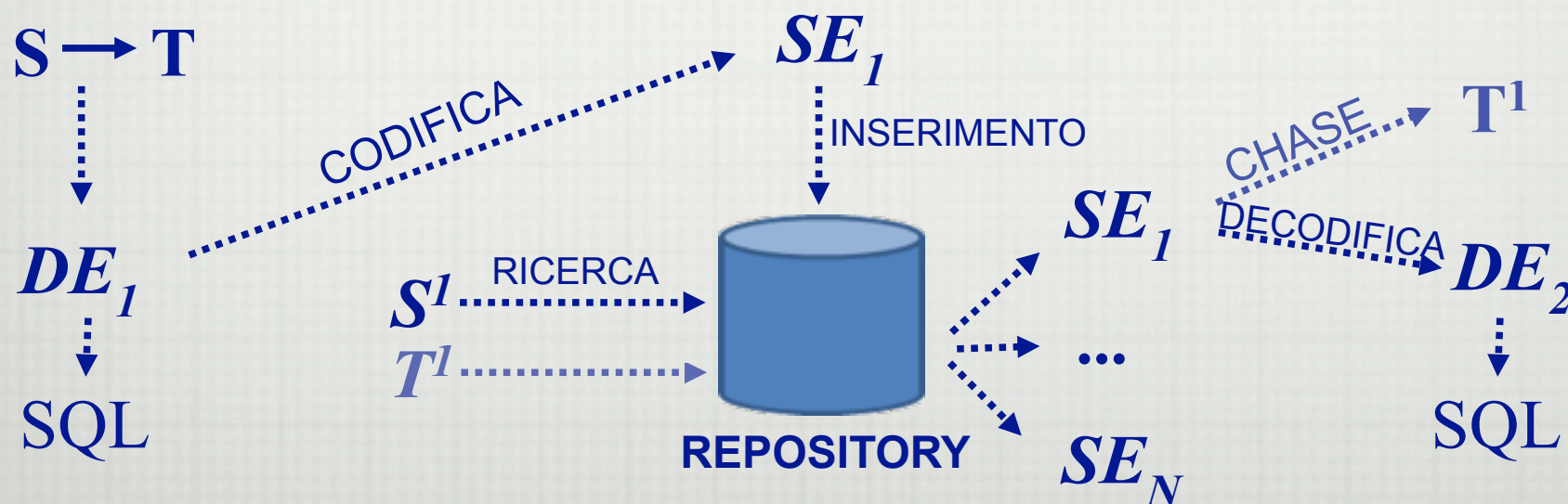
PROCESSO

1. DATO SCHEMA SOURCE S E SCHEMA TARGET T .
2. GENERA UN DATA EXCHANGE DE_1 CHE PRODUCE UNA TRASFORMAZIONE SUI DATI IN SQL
3. IL DATA EXCHANGE VIENE CODIFICATO IN UNO SCHEMA EXCHANGE SE_1 CHE VIENE INSERITO NEL REPOSITORY.
4. DATO UN NUOVO SCHEMA SOURCE S^1 DI STRUTTURA SIMILE AD S , GAIA RESTITUISCE GLI SCHEMA EXCHANGE PRESENTI NEL REPOSITORY ADATTI PER S^1 .



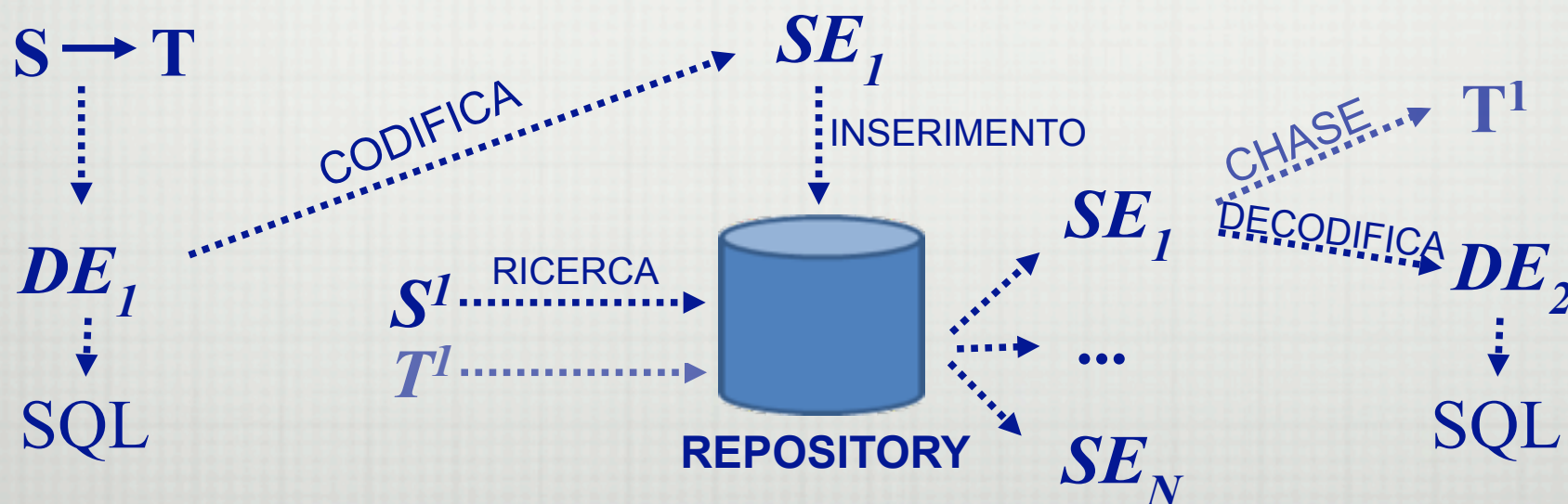
PROCESSO

1. DATO SCHEMA SOURCE S E SCHEMA TARGET T .
2. GENERA UN DATA EXCHANGE DE_1 CHE PRODUCE UNA TRASFORMAZIONE SUI DATI IN SQL
3. IL DATA EXCHANGE VIENE CODIFICATO IN UNO SCHEMA EXCHANGE SE_1 CHE VIENE INSERITO NEL REPOSITORY.
4. DATO UN NUOVO SCHEMA SOURCE S^1 DI STRUTTURA SIMILE AD S , GAIA RESTITUISCE GLI SCHEMA EXCHANGE PRESENTI NEL REPOSITORY ADATTI PER S^1 .
5. SE_1 (TRAMITE LA PROCEDURA DEL CHASE) PRODUCE LO SCHEMA TARGET T^1 .



PROCESSO

1. DATO SCHEMA SOURCE S E SCHEMA TARGET T .
2. GENERA UN DATA EXCHANGE DE_1 CHE PRODUCE UNA TRASFORMAZIONE SUI DATI IN SQL
3. IL DATA EXCHANGE VIENE CODIFICATO IN UNO SCHEMA EXCHANGE SE_1 CHE VIENE INSERITO NEL REPOSITORY.
4. DATO UN NUOVO SCHEMA SOURCE S^1 DI STRUTTURA SIMILE AD S , GAIA RESTITUISCE GLI SCHEMA EXCHANGE PRESENTI NEL REPOSITORY ADATTI PER S^1 .
5. SE_1 (TRAMITE LA PROCEDURA DEL CHASE) PRODUCE LO SCHEMA TARGET T^1 .
6. SE_1 VIENE DECODIFICATO E VA A FORMARE UN DATA EXCHANGE DE_2 CHE VIENE UTILIZZATO PER CREARE UNA NUOVA TRASFORMAZIONE SUI DATI.



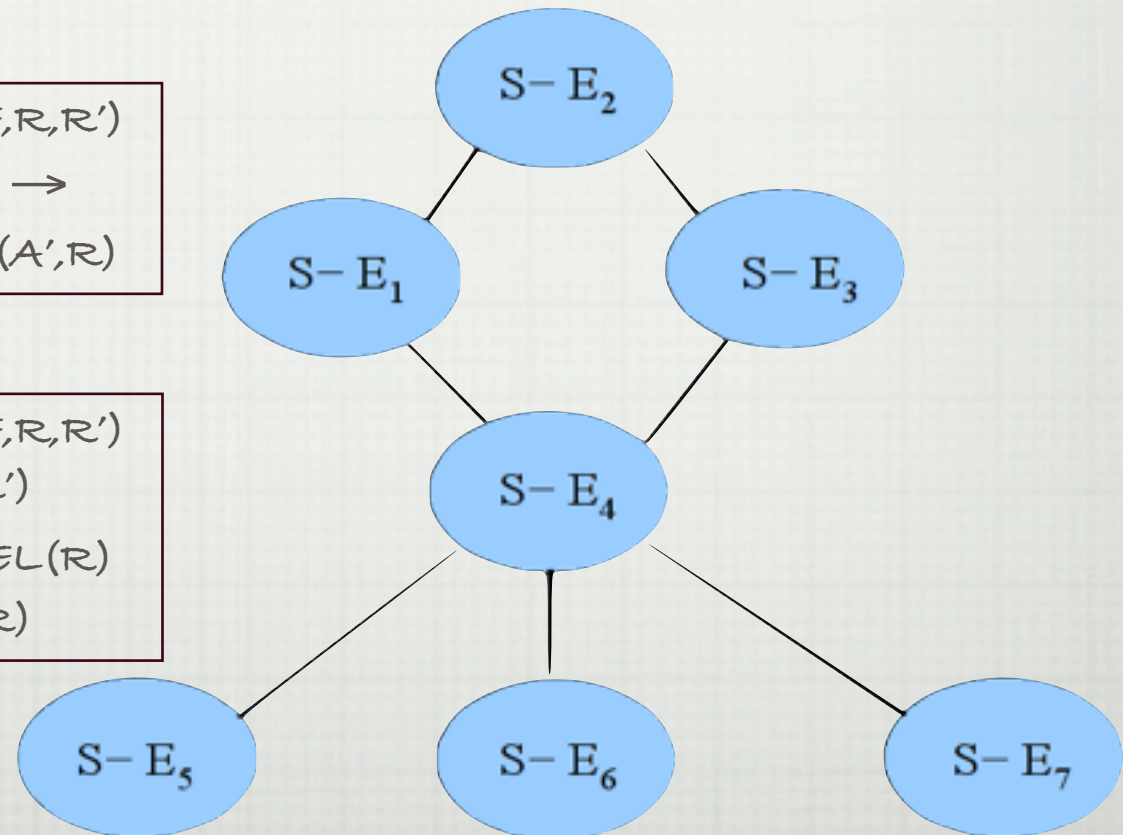
ORDINAMENTO DEI RISULTATI

GLI SCHEMA EXCHANGE RESTITUITI COME RISULTATO DALLA RICERCA NEL REPOSITORY VENGONO ORDINATI SECONDO UNA RELAZIONE ANTENATO-DISCENDENTE, IN CUI CIASCUN ANTENATO RAPPRESENTA UNA TRASFORMAZIONE PIÙ GENERALE RISPETTO A QUELLA DI TUTTI I SUOI DISCENDENTI.

REL(R) KEY(K,R) ATT(A,R) FK(F,R,R')
REL(R') KEY(K',R') ATT(A',R') →
REL(R) KEY(K,R) ATT(A,R) ATT(A',R)

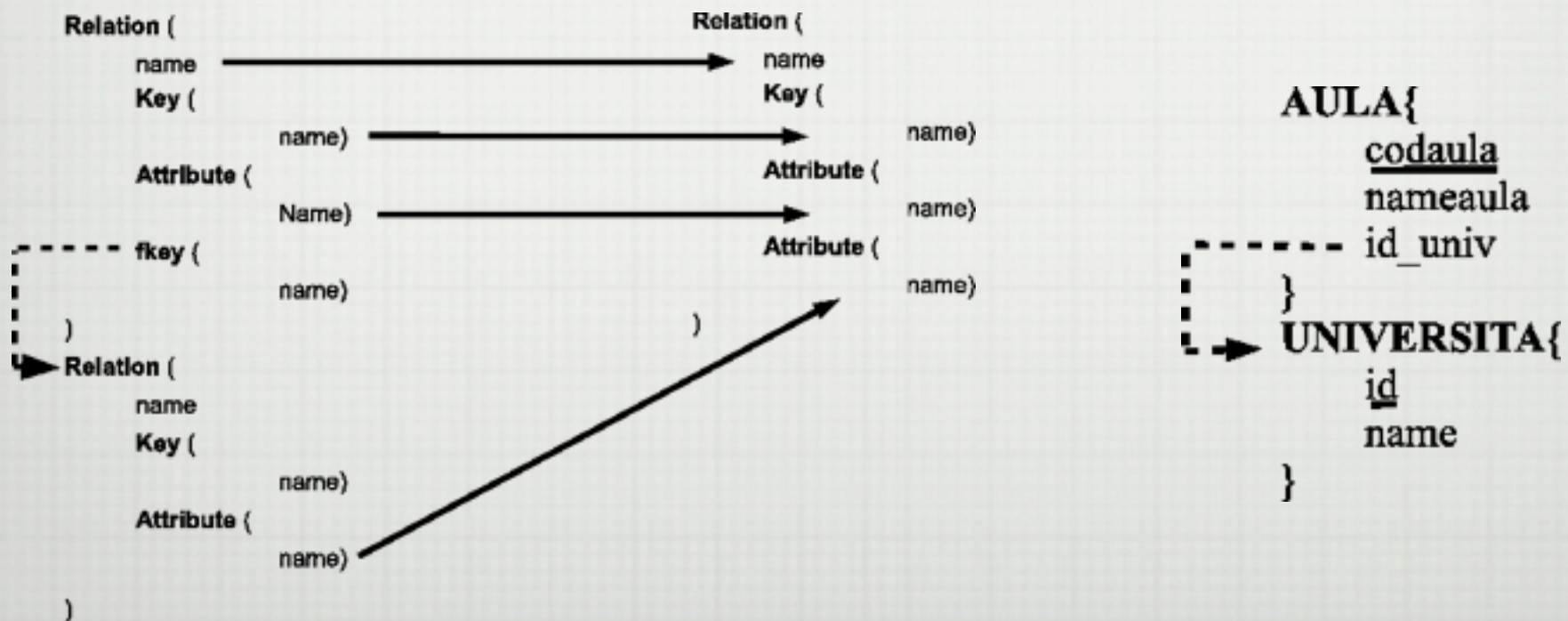
...

REL(R) KEY(K,R) ATT(A,R) FK(F,R,R')
REL(R') KEY(K',R') ATT(A',R')
(K='SSN') (A='PHONE #') → REL(R)
KEY(K,R) ATT(A,R) ATT(A',R)



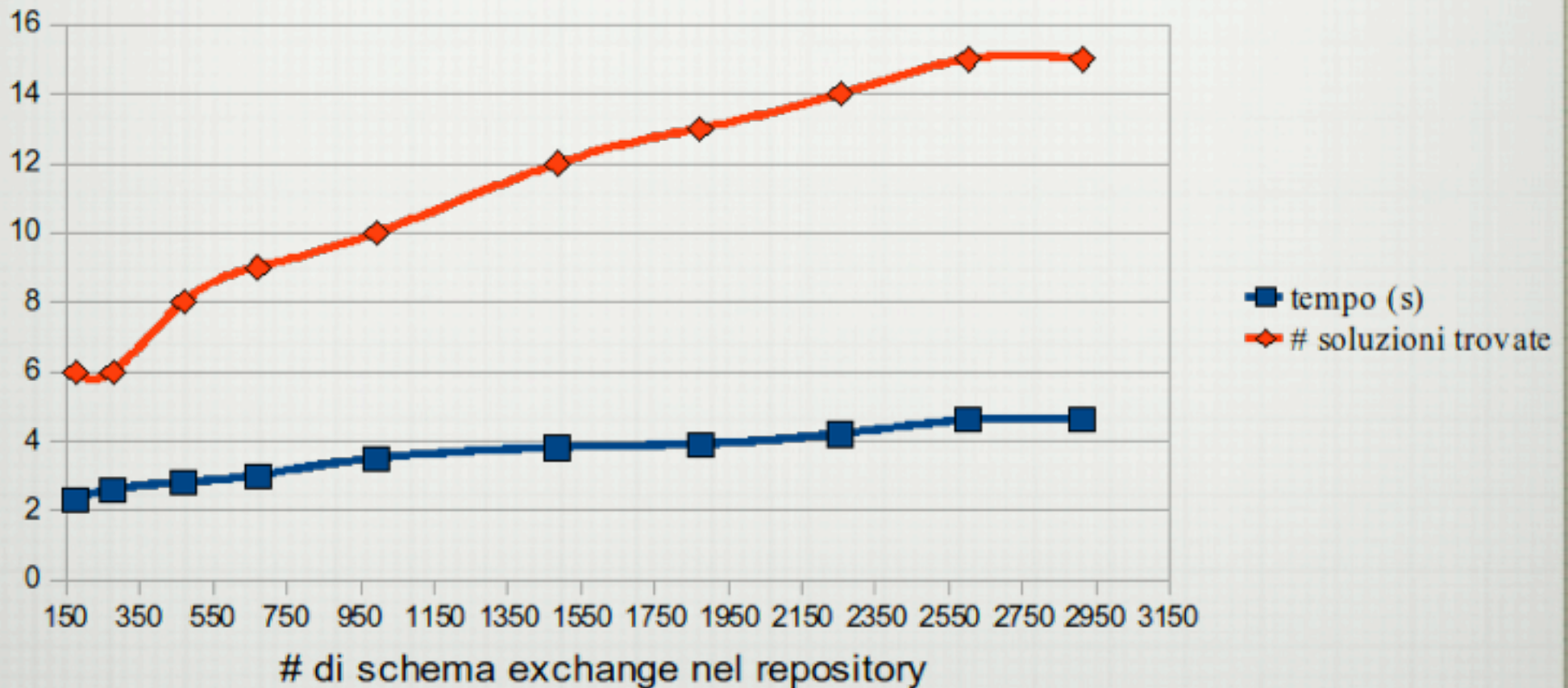
ESPERIMENTO SCALABILITÀ I

1. GENERAZIONE DI UN NUMERO CRESCENTE DI SCHEMA EXCHANGE CASUALI CON STRUTTURA BEN DEFINITA AGGIUNGENDO COSTANTI DI UGUAGLIANZA AI VARI ELEMENTI SCELTE RANDOMICAMENTE;
2. SI TENTA DI INSERIRLI NEL REPOSITORY; SE GIÀ PRESENTI, IL SISTEMA LI SCARTA;
3. SI CERCANO GLI SCHEMA EXCHANGE APPLICABILI AD UNO SCHEMA DI BASE DI DATI;
4. SI MONITORA, A FRONTE DEL NUMERO DI SCHEMA EXCHANGE PRESENTI NEL REPOSITORY, IL NUMERO DI SOLUZIONI TROVATE ED IL TEMPO DI ESECUZIONE DELLA RICERCA (È UN'ANALISI DI CASO PEGGIORE);



ESPERIMENTO SCALABILITÀ 2

COME SI COMPORTA LA RICERCA QUANDO CI SONO MOLTI DATI NEL REPOSITORY



PER MAGGIORI INFORMAZIONI

- [HTTP://GAIA.DIA.UNIROMA3.IT/](http://gaia.dia.uniroma3.it/)

