Tutorial 3: Relational Algebra and Tuple-Relational Calculus

Informatics 1 Data & Analysis

Week 5, Semester 2, 2013–2014

This worksheet has three parts: tutorial Questions, followed by some Examples and their Solutions.

- Before your tutorial, work through and attempt all of the Questions in the first section.
- The Examples are there for additional preparation, practice, and revision.
- Use the Solutions to check your answers, and read about possible alternatives.

You must bring your answers to the main questions along to your tutorial. You will need to be able to show these to your tutor, and may be exchanging them with other students, so it is best to have them printed out on paper.

If you cannot do some questions, write down what it is that you find challenging and use this to ask your tutor in the meeting.

Tutorials will not usually cover the Examples, but if you have any questions about those then write them down and ask your tutor, or go along to InfBASE during the week.

It's important both for your learning and other students in the group that you come to tutorials properly prepared. If you have not attempted the main tutorial questions, then you may be sent away from the tutorial to do them elsewhere.

Data & Analysis tutorials are not formally assessed, but they are a compulsory and important part of the course. If you do not do the exercises then you are unlikely to pass the exam.

Attendance at tutorials is obligatory: if you are ill or otherwise unable to attend one week then email your tutor, and if possible attend another tutorial group in the same week.

Optional Reading: Chapter 4 (Relational Algebra and Calculus) of Ramakrishnan and Gehrke, *Database Management Systems*; or the equivalent in any other database textbook (see Lecture 4 for six of these).

Please send any corrections and suggestions to Ian.Stark@ed.ac.uk

Introduction

In this tutorial, you will construct queries in *tuple-relational calculus* and describe operations to compute their results using *relational algebra*. These systems were introduced, with examples, in the lectures. All questions in this tutorial are based on a set of relational tables dealing with air travel: airports, flights, bookings and seats. You may find this tutorial bit more difficult than the previous ones. If you need any help, please ask the drop-in lab demonstrators or go to InfBASE for assistance. If you are stuck with any question, write down what you are finding difficult and then move on to try the next one.

A Relational Model for Flight Bookings

The following DDL declarations and table data describe a relational model for air travel bookings. This is a very simplified model — in particular, we don't deal with times or dates of individual flights. Notice that we are taking advantage of SQL's case insensitivity for keywords, with **create table** and **primary key** instead of **CREATE TABLE** and **PRIMARY KEY**.

```
create table Airport (
    airportId  varchar(3),
    name       varchar(50),
    city       varchar(30),
    primary key (airportId)
)

create table Flight (
    flightNo         varchar(6),
    flightCompany    varchar(20),
    depAirport       varchar(3),
    arrAirport       varchar(3),
    primary key (flightNo),
    foreign key (depAirport) references Airport(airportId),
    foreign key (arrAirport)  references Airport(airportId)
)

create table Booking (
    ticketNo     varchar(9),
    name         varchar(20),
    nationality  varchar(20),
    flightNo     varchar(6),
    seatNo       varchar(3),
    primary key (ticketNo),
    foreign key (flightNo) references Flight,
    foreign key (seatNo, flightNo) references Seat
)

create table Seat (
    seatNo    varchar(3),
    flightNo  varchar(6),
    class     varchar(10),
    primary key (seatNo, flightNo),
    foreign key (flightNo) references Flight
)
```

Airport

airportId	name	city
LHR	Heathrow	London
LGW	Gatwick	London
CDG	Charles de Gaulle	Paris
ORY	Orly	Paris

Flight

flightNo	flightCompany	depAirport	arrAirport
AF1231	Air France	LHR	CDG
AF1232	Air France	CDG	LHR
AF1234	Air France	LGW	CDG
AF1235	Air France	CDG	LGW
BA2943	British Airways	LGW	ORY
BA2944	British Airways	ORY	LGW
BA4059	British Airways	LHR	CDG
BA4060	British Airways	CDG	LHR

Booking

ticketNo	name	nationality	flightNo	seatNo
EAG129489	John Jones	British	AF1232	12D
EAF123456	Fraser McEwan	British	AF1232	30E
ABS958332	Jane Smith	French	BA2944	10A
ORE394895	Fiona Stewart	British	BA4060	5D
EYR149583	Tom Woods	British	BA4059	14B
EAG348595	John Smith	French	BA2944	30D

Seat

seatNo	flightNo	class
12D	AF1232	Business
30E	AF1232	Economy
10A	BA2944	Business
5D	BA4060	Business
14B	BA4059	Economy
30D	BA2944	Economy

Question 1: Operations in Relational Algebra

For each of the following queries in relational algebra, calculate the output table and give a brief statement of what query it answers.

(a) $\sigma_{\mathsf{class='Business'}}(\mathsf{Seat})$

(b) $\pi_{\mathsf{nationality}}(\mathsf{Booking})$

(c) $\sigma_{\mathsf{nationality='French'}}(\mathsf{Booking}) \times \sigma_{\mathsf{class='Business'}}(\mathsf{Seat})$

(d) $\mathsf{Booking} \bowtie \mathsf{Seat}$

(e) $\pi_{\mathsf{name}}(\sigma_{\mathsf{class='Business'}}(\mathsf{Booking} \bowtie \mathsf{Seat}))$

(f) $\mathsf{Airport} \cup \mathsf{Seat}$

Question 2: Constructing Queries

For each of the following questions, formulate the specified queries in tuple-relational calculus and as a computation in relational algebra.

(a) Retrieve all information about airports in London. The schema of the output table should be same as that of the Airport table.

(b) Retrieve details of all bookings by British and French passengers. The schema of the output table should be same as that of the Booking table.

(c) Retrieve the names of all passengers.

(d) Retrieve the flight number, Departure and Arrival airports of all British Airways flights.

(e) Retrieve the name of every passenger together with their flight number and the associated flight company.

The following questions are all marked with a star ⋆. This indicates that they are optional — you are encouraged to attempt all you can, but they are not a requirement for tutorials.

⋆ **(f)** Retrieve details of all flights from all airports in London. The output schema should be same as that of Flight table.

⋆ **(g)** Find out the ticket numbers and names of all passengers departing from London.

⋆ **(h)** Retrieve the flight number and company of all flights from London to Paris.

Examples

This section contains further exercises on constructing queries in tuple-relational calculus and relational algebra. These examples are similar to the main tutorial questions: A relational model for a given domain is provided, and queries in the two formalisms are defined.

Following these there is a section presenting solutions and notes on all the examples.

A Relational Model for Films

The following DDL declarations and table data describe a relational model for the film industry.

```
create table Actor (
    actorId         varchar(5),
    name            varchar(50),
    nationality     varchar(20),
    primary key (actorId)
)

create table Film (
    filmId          varchar(5),
    title           varchar(50),
    year            integer,
    directorId      varchar(5),
    primary key (filmId),
    foreign key (directorId)  references Director
)

create table PerformsIn (
    actorId         varchar(5),
    filmId          varchar(5),
    character       varchar(50),
    primary key (actorId, filmId),
    foreign key (actorId)  references Actor,
    foreign key (filmId)   references Film
)

create table Director (
    directorId      varchar(5),
    name            varchar(50),
    nationality     varchar(20),
    primary key (directorId)
)
```

Actor

actorId	name	nationality	age
LDC21	Leonardo DiCaprio	American	40
KW871	Kate Winslet	British	39
CB379	Christian Bale	British	40
MKE12	Michael Keaton	American	63
JGL81	Joseph Gordon-Levitt	American	33
EMG32	Ewan McGregor	British	43
HBC54	Helena Bonham Carter	British	48

Film

filmId	title	year	directorId
INC10	Inception	2010	CN345
TIT97	Titanic	1997	JC212
RR008	Revolutionary Road	2008	SM521
SKF12	Skyfall	2012	SM521
SHI10	Shutter Island	2010	SCO78
DK008	The Dark Knight	2008	CN345
DKR12	The Dark Knight Rises	2012	CN345
BAT92	Batman Returns	1992	BUR34
FISH4	Big Fish	2003	BUR34

PerformsIn

actorId	filmId	character
LDC21	INC10	Dominic Cobb
LDC21	TIT97	Jack Dawson
KW871	TIT97	Rose DeWitt Bukater
LDC21	RR008	Frank Wheeler
KW871	RR008	April Wheeler
LDC21	SHI10	Teddy Daniels
CB379	DK008	Bruce Wayne
CB379	DKR12	Bruce Wayne
JGL81	INC10	Arthur
MKE12	BAT92	Bruce Wayne
EMG32	FISH4	Ed Bloom
HBC54	FISH4	Jenny

Director

directorId	name	nationality
CN345	Christopher Nolan	British
JC212	James Cameron	Canadian
SM521	Sam Mendes	British
SCO78	Martin Scorsese	American
BUR34	Tim Burton	American

Example 1: Operations in Relational Algebra

For each of the following queries in relational algebra, calculate the output table and give a brief statement of what query it answers.

(a) $\sigma_{\mathsf{age}>45}(\mathsf{Actor})$

(b) $\pi_{\mathsf{title}}(\mathsf{Film})$

(c) $\pi_{\mathsf{title}}(\sigma_{\mathbf{year}<2000}(\mathsf{Film}))$

(d) $\sigma_{\mathbf{year}=2012}(\mathsf{Film}) \times \sigma_{\mathsf{nationality}\neq\text{'American'}}(\mathsf{Director})$

(e) $\sigma_{\mathbf{year}=2012}(\mathsf{Film}) \bowtie \sigma_{\mathsf{nationality}\neq\text{'American'}}(\mathsf{Director})$

(f) $\pi_{\mathsf{title}}(\mathsf{Film} \bowtie \sigma_{\mathsf{nationality}=\text{'British'}}(\mathsf{Director}))$

(g) $\sigma_{\mathbf{year}<2000}(\mathsf{Film}) \cup \sigma_{\mathbf{year}>2010}(\mathsf{Film})$

(h) $\sigma_{\mathbf{year}\geq2000}(\mathsf{Film}) \cap \sigma_{\mathbf{year}\leq2010}(\mathsf{Film})$

Example 2: Constructing Queries

For each of the following questions, formulate the specified queries in tuple-relational calculus and as a computation in relational algebra.

(a) Retrieve details of all films that were released in 2010. The output schema should be the same as that of the Film table.

(b) Retrieve details of all actors that are not in their thirties. The output schema should be the same as that of the Actor table.

(c) Retrieve the names of all directors.

(d) Retrieve the names of all American directors.

(e) Find out the names of all British actors above the age of 40.

(f) Retrieve the name of each actor together with the titles of the films he/she has performed in.

(g) Find out the names of all actors that have played the character of Bruce Wayne (Batman).

(h) Retrieve the names of all actors that have played the character of Bruce Wayne, together with the year the corresponding films were released.

(i) Retrieve all actors from the film Inception. The output schema should be the same as that of the Actor table.

(j) Find out the names of all actors that have performed in a film directed by Christopher Nolan.

(k) Retrieve the titles of all films in which Leonardo Di Caprio and Kate Winslet have co-acted.

(l) Assuming that the ids of actors and directors are used consistently across the tables, retrieve details of all actors that have directed a film.

Solutions to Examples

These are not entirely "model" answers; instead, they indicate a possible solution. Remember that not all of these questions will have a single "right" answer. There can be multiple appropriate ways to formulate a query.

If you have difficulties with a particular example, or have trouble following through the solution, please raise this as a question in your tutorial.

Solution 1

(a) $\sigma_{\mathsf{age}>45}(\mathsf{Actor})$

Retrieves details of all actors above the age of 45. The output table is as follows:

actorId	name	nationality	age
MKE12	Michael Keaton	American	63
HBC54	Helena Bonham Carter	British	48

(b) $\pi_{\mathsf{title}}(\mathsf{Film})$

Retrieves all distinct film titles. The output table is as follows:

title
Inception
Titanic
Revolutionary Road
Skyfall
Shutter Island
The Dark Knight
The Dark Knight Rises
Batman Returns
Big Fish

(c) $\pi_{\mathsf{title}}(\sigma_{\mathsf{year}<2000}(\mathsf{Film}))$

Retrieves all distinct titles of films that were released before 2000. The output table is as follows:

title
Titanic
Batman Returns

(d) $\sigma_{\mathsf{year}=2012}(\mathsf{Film}) \times \sigma_{\mathsf{nationality}\neq\text{'American'}}(\mathsf{Director})$

Retrieves all information about all combinations of films released in 2012 and non-American directors. The output table is as follows:

filmId	title	year	directorId	directorId	name	nationality
SKF12	Skyfall	2012	SM521	CN345	Christopher Nolan	British
SKF12	Skyfall	2012	SM521	JC212	James Cameron	Canadian
SKF12	Skyfall	2012	SM521	SM521	Sam Mendes	British
DKR12	The Dark Knight Rises	2012	CN345	CN345	Christopher Nolan	British
DKR12	The Dark Knight Rises	2012	CN345	JC212	James Cameron	Canadian
DKR12	The Dark Knight Rises	2012	CN345	SM521	Sam Mendes	British

(e) $\sigma_{\textsf{year}=2012}(\textsf{Film}) \bowtie \sigma_{\textsf{nationality}\neq\textsf{'American'}}(\textsf{Director})$

Retrieves the details of all films released in 2012 and directed by a non-American director, along with the details of the corresponding director. The output table is as follows:

filmId	title	year	directorId	name	nationality
SKF12	Skyfall	2012	SM521	Sam Mendes	British
DKR12	The Dark Knight Rises	2012	CN345	Christopher Nolan	British

(f) $\pi_{\textsf{title}}(\textsf{Film} \bowtie \sigma_{\textsf{nationality}=\textsf{'British'}}(\textsf{Director}))$

Retrieves all distinct titles of films directed by a British director. The output table is as follows:

title
Inception
Revolutionary Road
Skyfall
The Dark Knight
The Dark Knight Rises

(g) $\sigma_{\textsf{year}<2000}(\textsf{Film}) \cup \sigma_{\textsf{year}>2010}(\textsf{Film})$

Retrieves details of all films released before 2000 or after 2010. The output table is as follows:

filmId	title	year	directorId
TIT97	Titanic	1997	JC212
SKF12	Skyfall	2012	SM521
DKR12	The Dark Knight Rises	2012	CN345
BAT92	Batman Returns	1992	BUR34

(h) $\sigma_{\textsf{year}\geq2000}(\textsf{Film}) \cap \sigma_{\textsf{year}\leq2010}(\textsf{Film})$

Retrieves details of all films released between 2000 and 2010. The output table is as follows:

filmId	title	year	directorId
INC10	Inception	2010	CN345
RR008	Revolutionary Road	2008	SM521
SHI10	Shutter Island	2010	SCO78
DK008	The Dark Knight	2008	CN345
FISH4	Big Fish	2003	BUR34

Solution 2

(a) Retrieve details of all films that were released in 2010. The output schema should be the same as that of the Film table.

$$\{F \mid F \in \textsf{Film} \wedge F.\textsf{year} = 2010\ \}$$
$$\sigma_{\textsf{year}=2010}(\textsf{Film})$$

(b) Retrieve details of all actors that are not in their thirties. The output schema should be the same as that of the Film table.

$$\{A \mid A \in \textsf{Actor} \wedge (A.\textsf{age} < 30 \vee A.\textsf{age} > 39)\ \}$$
$$\sigma_{(\textsf{age}<30\vee\textsf{age}>39)}(\textsf{Actor})$$
$$\text{or} \quad \sigma_{\textsf{age}<30}(\textsf{Actor}) \cup \sigma_{\textsf{age}>39}(\textsf{Actor})$$

(c) Retrieve the names of all directors.

$$\{\ T \mid \exists D \in \mathsf{Director}\ .\ T.\mathsf{name} = D.\mathsf{name}\ \}$$
$$\pi_{\mathsf{name}}(\mathsf{Director})$$

(d) Retrieve the names of all American directors.

$$\{\ T \mid \exists D \in \mathsf{Director}\ .\ D.\mathsf{nationality} = \text{'American'} \wedge T.\mathsf{name} = D.\mathsf{name}\ \}$$
$$\pi_{\mathsf{name}}(\sigma_{\mathsf{nationality}=\text{'American'}}(\mathsf{Director}))$$

(e) Find out the names of all British actors above the age of 40.

$$\{\ T \mid \exists A \in \mathsf{Actor}\ .\ A.\mathsf{nationality} = \text{'British'} \wedge A.\mathsf{age} > 40 \wedge T.\mathsf{name} = A.\mathsf{name}\ \}$$
$$\pi_{\mathsf{name}}(\sigma_{(\mathsf{nationality}=\text{'British'} \wedge \mathsf{age}>40)}(\mathsf{Actor}))$$

(f) Retrieve the name of each actor together with the titles of the films he/she has performed in.

$$\{T\ \mid \exists A \in \mathsf{Actor}, P \in \mathsf{PerformsIn}, F \in \mathsf{Film}\ .\ A.\mathsf{actorId} = P.\mathsf{actorId} \wedge P.\mathsf{filmId} = F.\mathsf{filmId} \wedge T.\mathsf{name} = A.\mathsf{name} \wedge T.\,\mathsf{title}\ = F.\,\mathsf{title}\ \}$$
$$\pi_{\mathsf{name},\,\mathsf{title}}\,(\mathsf{Actor} \bowtie (\mathsf{PerformsIn} \bowtie \mathsf{Film}))$$

(g) Find out the names of all actors that have played the character of Bruce Wayne (Batman; see Marshall et al., Physics Special Topics 10(1):2011).

$$\{T\ \mid \exists A \in \mathsf{Actor}, P \in \mathsf{PerformsIn}\ .\ A.\mathsf{actorId} = P.\mathsf{actorId} \wedge P.\mathsf{character} = \text{'Bruce Wayne'} \wedge T.\mathsf{name} = A.\mathsf{name}\ \}$$
$$\pi_{\mathsf{name}}(\mathsf{Actor} \bowtie (\sigma_{\mathsf{character}=\text{'Bruce Wayne'}}(\mathsf{PerformsIn}))$$

(h) Retrieve the names of all actors that have played the character of Bruce Wayne, together with the year the corresponding films were released.

$$\{T\ \mid \exists A \in \mathsf{Actor}, P \in \mathsf{PerformsIn}, F \in \mathsf{Film}\ .\ A.\mathsf{actorId} = P.\mathsf{actorId} \wedge P.\mathsf{filmId} = F.\mathsf{filmId} \wedge P.\mathsf{character} = \text{'Bruce Wayne'} \wedge T.\mathsf{name} = A.\mathsf{name} \wedge T.\mathbf{year} = F.\mathbf{year}\ \}$$
$$\pi_{\mathsf{name},\mathbf{year}}(\mathsf{Actor} \bowtie (\sigma_{\mathsf{character}=\text{'Bruce Wayne'}}(\mathsf{PerformsIn} \bowtie \mathsf{Film})))$$

(i) Retrieve all actors that appeared in Inception. The output schema should be the same as that of the Actor table.

$$\{A\ \mid A \in \mathsf{Actor}, \exists P \in \mathsf{PerformsIn}, F \in \mathsf{Film}\ .\ A.\mathsf{actorId} = P.\mathsf{actorId} \wedge P.\mathsf{filmId} = F.\mathsf{filmId} \wedge F.\,\mathsf{title}\ = \text{'Inception'}\ \}$$
$$\pi_{\mathsf{actorId},\mathsf{name},\mathsf{nationality},\mathsf{age}}(\mathsf{Actor} \bowtie (\mathsf{PerformsIn} \bowtie (\sigma_{\,\mathsf{title}\,=\text{'Inception'}}(\mathsf{Film}))))$$

(j) Find out the names of all actors that have performed in a film directed by Christopher Nolan.

$$\{T\ \mid \exists A \in \mathsf{Actor}, P \in \mathsf{PerformsIn}, F \in \mathsf{Film}, D \in \mathsf{Director}\ .\ A.\mathsf{actorId} = P.\mathsf{actorId} \wedge P.\mathsf{filmId} = F.\mathsf{filmId} \wedge F.\mathsf{directorId} = D.\mathsf{directorId} \wedge D.\mathsf{name} = \text{'Christopher Nolan'} \wedge T.\mathsf{name} = A.\mathsf{name}\ \}$$
$$\pi_{\mathsf{name}}(\mathsf{Actor} \bowtie \mathsf{PerformsIn} \bowtie \mathsf{Film} \bowtie (\sigma_{\mathsf{name}=\text{'Christopher Nolan'}}(\mathsf{Director})))$$

(k) Retrieve the titles of all films in which Leonardo Di Caprio and Kate Winslet have co-acted.

$$\{T\ \mid \exists A1 \in \mathsf{Actor}, A2 \in \mathsf{Actor}, P1 \in \mathsf{PerformsIn}, P2 \in \mathsf{PerformsIn}, F \in \mathsf{Film}\ .\ A1.\mathsf{actorId} = P1.\mathsf{actorId} \wedge A2.\mathsf{actorId} = P2.\mathsf{actorId} \wedge A1.\mathsf{name} = \text{'Leonardo DiCaprio'} \wedge A2.\mathsf{name} = \text{'Kate Winslet'} \wedge P1.\mathsf{filmId} = P2.\mathsf{filmId} \wedge F.\mathsf{filmId} = P1.\mathsf{filmId} \wedge T.\,\mathsf{title}\ = F.\,\mathsf{title}\ \}$$
$$\pi_{\,\mathsf{title}}\,((\pi_{\mathsf{filmId}}(\sigma_{\mathsf{name}=\text{'Leonardo DiCaprio'}}(\mathsf{Actor}) \bowtie \mathsf{PerformsIn}) \cap \pi_{\mathsf{filmId}}(\sigma_{\mathsf{name}=\text{'Kate Winslet'}}(\mathsf{Actor}) \bowtie \mathsf{PerformsIn})) \bowtie \mathsf{Film})$$

(l) Assuming that the ids of actors and directors are consistent across the tables, retrieve details of all actors that have directed a film.

$$\{A \mid A \in \mathsf{Actor}, \exists D \in \mathsf{Director} \;.\; A.\mathsf{actorId} = D.\mathsf{directorId}\ \}$$

$$\pi_{\mathsf{actorId,name,nationality,age}}(\mathsf{Actor} \bowtie_{\mathsf{actorId=directorId}} \mathsf{Director})$$