# Technologies for Information Systems

January 25, 2011
Duration: 1h45min

Prof. F.A. Schreiber
Prof. L. Tanca

Name: _____

Surname: _____

Student ID: _____

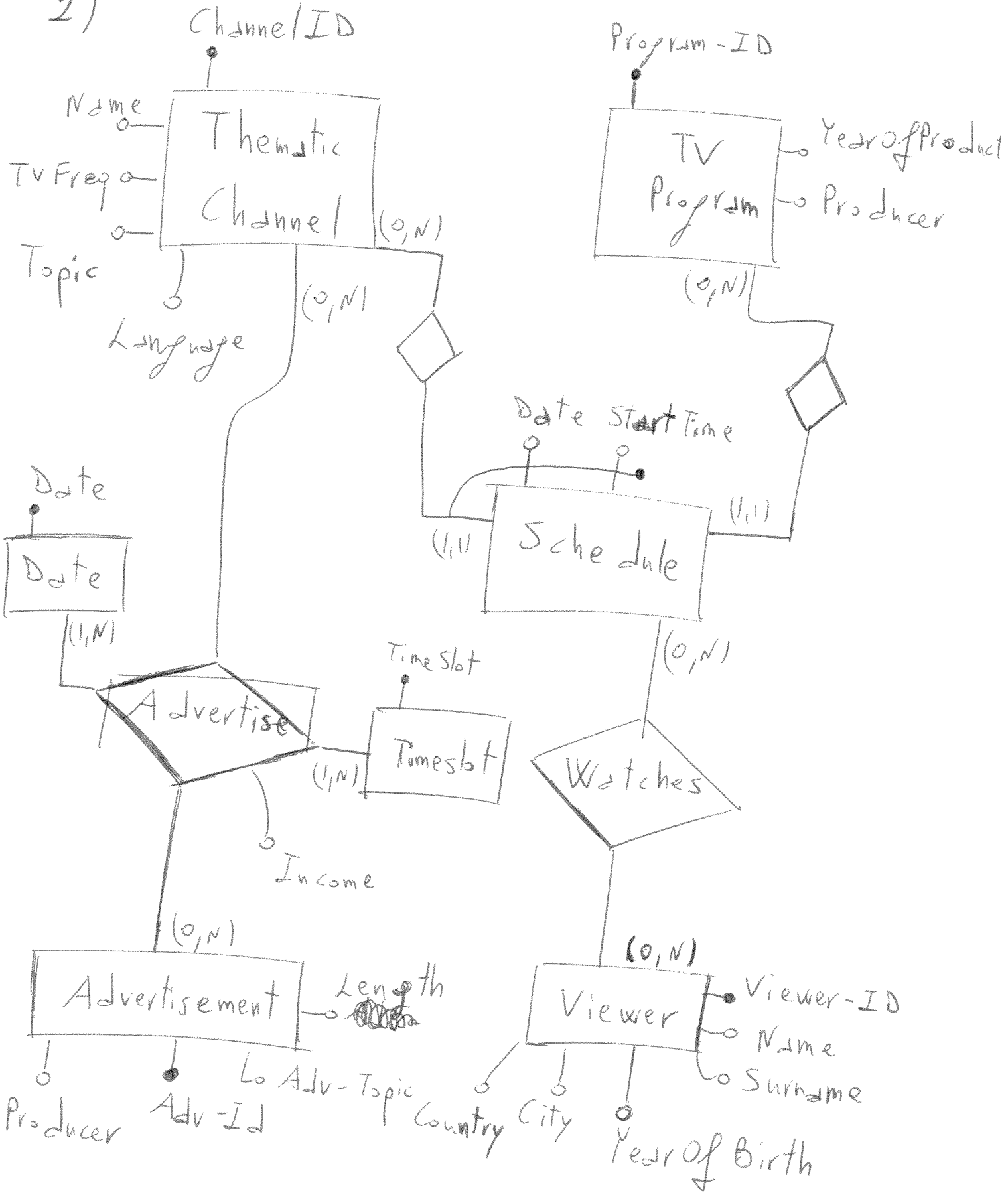
Signature: _____

PoliTV is a television network company and it has many thematic channels. The management of PoliTV has asked you to design its data warehouse in order to be able to analyze the audience ratings (e.g., number of viewers) and the advertising income. The following is the logical schema of the PoliTV operational database:

```
ThematicChannel(Channel-ID, Name, TVFrequency, Topic, Language)
TVProgram(Program-ID, YearOfProduction, Producer)
Advertisement(Adv-ID, Adv-Topic, Producer, Length)
Advertise(Channel-ID, Adv-ID, Date, Timeslot, Income)
Viewer(Viewer-ID, Name, Surname, YearOfBirth, City, Country)
Schedule(Channel-ID, Date, StartTime, Program-ID)
Watches(Viewer-ID, Channel-ID, Date, StartTime)
```

1. Perform the reverse engineering of the given logical schema into a conceptual schema.
2. With respect to the produced conceptual schema:
   (a) Discover the fact(s) that are useful to monitor PoliTV's activities. For each of them:
      i. Identify measures (with their glossary) and dimensions (with their hierarchies) and produce the attribute tree (with pruning and grafting).
      ii. Produce the conceptual schema (fact schema).
   (b) Produce a star schema or snowflake schema (with related discussion/motivation) consistent with the conceptual schema and such that it allows the following queries:
      i. select the total advertising income for each day of the week
      ii. considering only the first semester of year 2009 and only sport channels, select the total number of advertisements and the total income for each pair (channel, timeslot)
      iii. select the timeslot with the highest income
      iv. for each channel select the program with the highest audience (i.e., number of viewers)
   (c) Write the above queries in SQL.

1)


Channel ID
Name
Thematic Channel
TV Freq
Topic
Language
(0,N)
(0,N)
(0,N)
Program-ID
TV Program
Year of Product
Producer
(0,N)
Date
Start Time
(1,1)
(1,1)
Schedule
Date
Date
(1,N)
(0,N)
Time Slot
Advertise
Timeslot
(1,N)
Watches
Income
(0,N)
(0,N)
Advertisement
Length
Viewer
Viewer-ID
Name
Surname
Producer
Adv-Id
Adv-Topic
Country
City
Year of Birth

2.a) There are two facts

Fact: Advertising

Measures: Income
Number Of Advertisements

Dimensions: Channel,
Date,
Timeslot,
Advertisement

Glossary

Income.
Select SUM(Income)
From Advertise
Group by Channel-ID, Adv-Id,
Date, Timeslot;

Number of Advertisements.
Select COUNT(*)
From Advertise
Group by Channel-ID, Adv-Id,
Date, Timeslot;

Fact: Audience Ratings

Measures: Number of Viewers

Dimensions: Viewer's characteristic
TV Program,
Channel,
Date,
Time

Glossary

Number of Viewers
Select COUNT(*)
From Watches, Schedule, Viewer
Where - join conditions.
Group by Viewer.City,
Viewer.Year of Birth,
Channel-ID, Date, StartTime,
Program-ID;

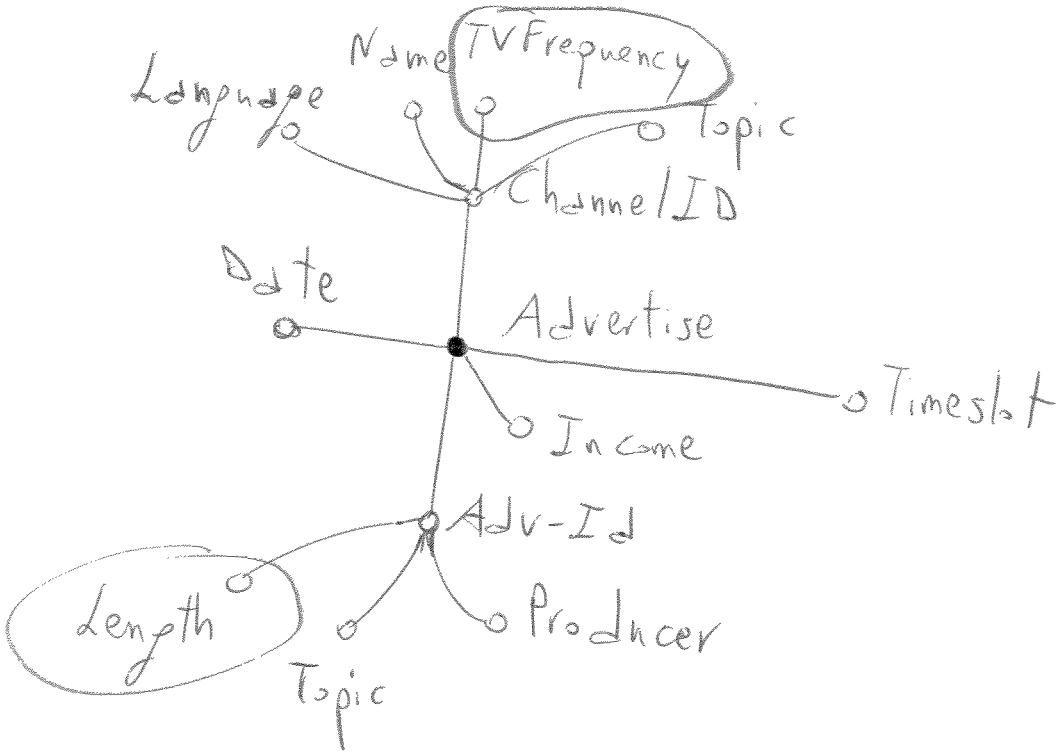
Discretize

# Attribute trees

 = pruning

 = grafting

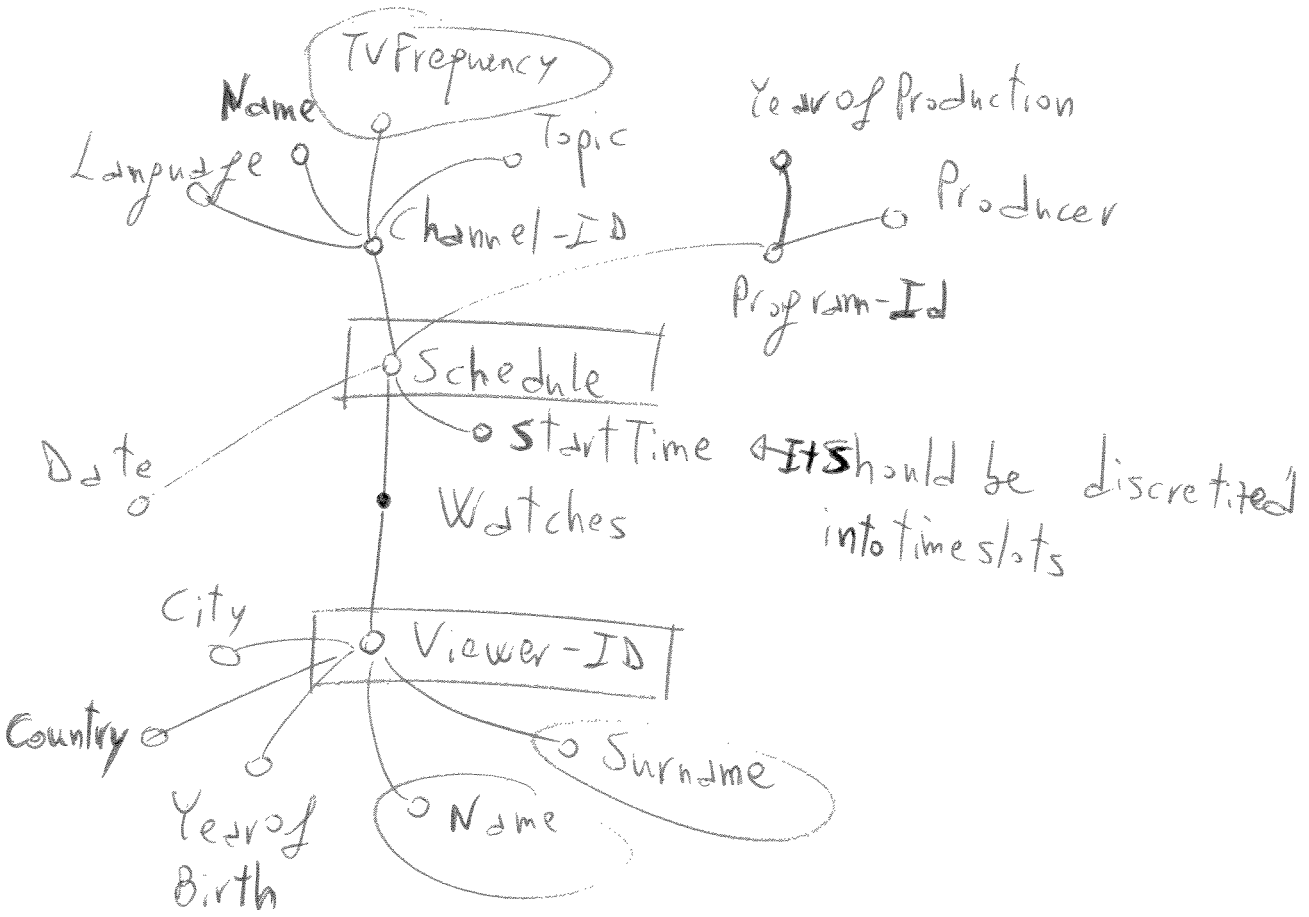
## Advertising



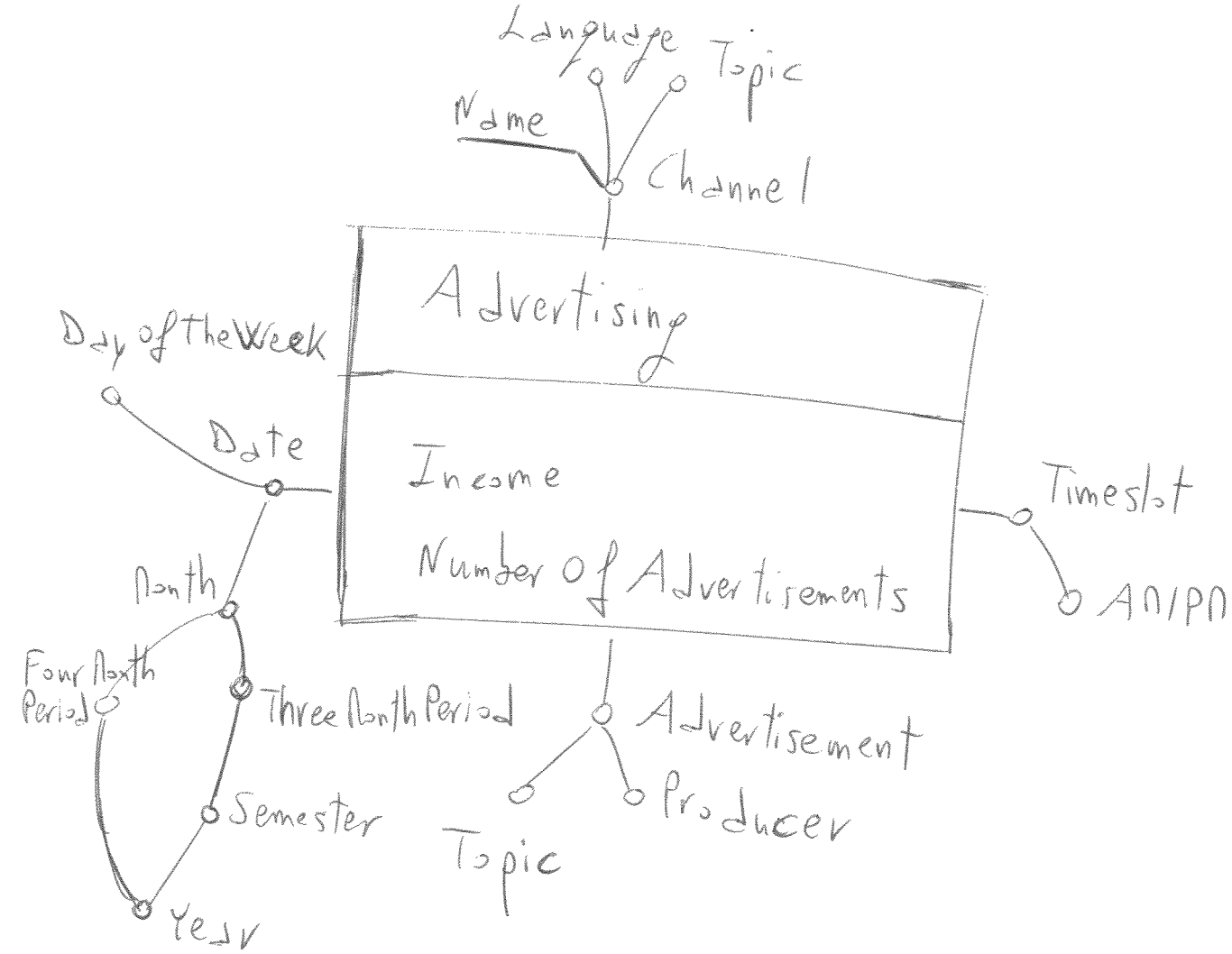
= pruning

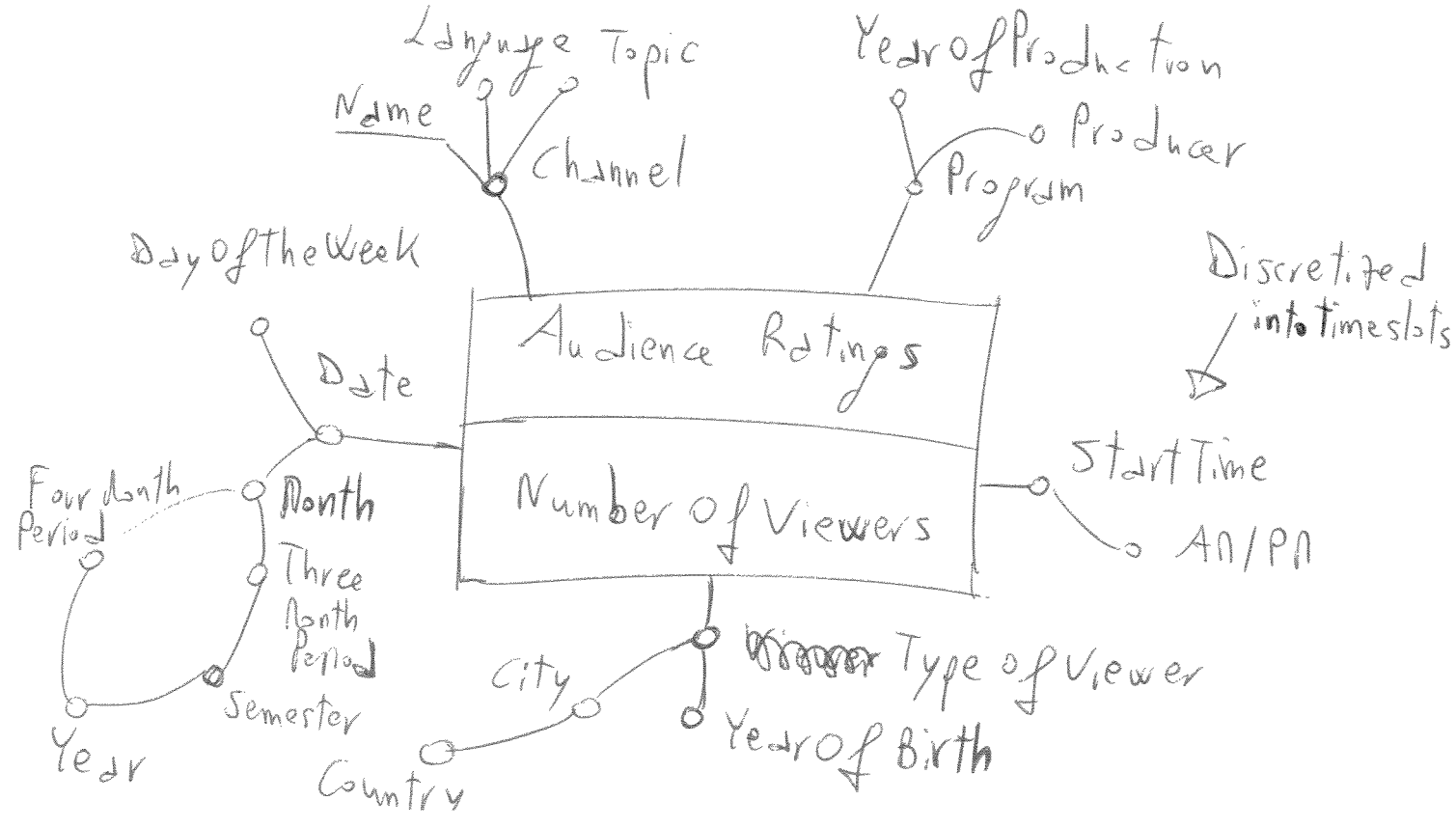
= grafting

# Audience Ratings



# Fact schema / Fact Model



2.b) Logical schema

Some dimensions are shared by the two facts.

Shared dimensions: Date, Channel.
Also "Time" can be shared if StartTime is discretized in timeslots.

FactAdvertising( ChannelID, DateID, AdvID, TimeID, Income, NumOfAdvs )

DimChannel ( ChannelID, Name, Language, Topic )

DimDate ( DateID, DayOfWeek, Month, ThreeMonthPeriod, FourMonthPeriod, Semester, Year )

DimAdvertisement ( AdvID, Topic, Producer )

DimeTime ( TimeID, Timeslot, AM_PM )

FactAudience ( ChannelID, ProgramID, TimeID, DateID, ViewerTypeID, NumOfViewers )

Dim Program ( Program ID , Producer, Year of Production)

Dim Type Viewer ( Viewer Type ID , Year of Birth, City, Country)

2.c) Sql queries

(i) Select SUM(Income), Day of Week
From Fact Advertising FA, Dim Date DD
where FA.DateID = DD.DateID
group by Day of Week;

(ii) Select SUM(Num of Advs), SUM(Income),
DC.ChannelID, DC.Name, DT.Timeslot
From Fact Advertising FA, Dim Time DT, Dim Channel DC,
Dim Date DD
where FA.TimeID = DT.TimeID AND
FA.ChannelID = DC.channelID AND
FA.DateID = DD.DateID AND
DD.Year = 2009 AND DD.Semester = 1 AND
DC.Topic = 'sport'
Group by DC.ChannelID, DC.Name, DT.TimeSlot;

(iii) Create view TimeslotIncome (TimeID, TotalIncome)
As
Select TimeID, SUM(Income)
From FactAdvertising
Group by TimeID;

Select DT.Timeslot
From DimTime, DT, TimeSlotIncome TS,
where DT.TimeID = TS.TimeID
and TotalIncome = (Select MAX(TotalIncome)
From TimeSlotIncome);

(iv) Create view ChannelProgramsAudience (ChannelID,
ProgramID, TotalAudience) AS
Select ChannelID, ProgramID,
SUM(NumofViewers)
From Fact Audience
Group by ChannelID, ProgramID;
Select DC.ChannelID, DC.Name, ProgramID
From ChannelProgramsAudience CPA1, DimChannel DC
where CPA1.ChannelID = DC.ChannelID AND
TotalAudience = (Select MAX(TotalAudience)

```
From   ChannelPrograms Audience
       CPA2
Where  CPA2.ChannelID =
       CPA1.ChannelID )
```

→ Another possible solution ( the former one is more efficient )

Create view ChannelPrograms Audience( - - )

=

```
Select DC.ChannelID, DC.Name, ProgramID
From ChannelPrograms Audience CPA1, DimChannel DC
Where CPA1.ChannelID = DC.ChannelID
and CPA1.ChannelID, [illegible] TotalAudience
  IN ( Select ChannelID, [illegible] MAX(TotalAudienc
       From ChannelPrograms Audience
       Group by ChannelID )
```