Technologies for Information Systems

January 25, 2011

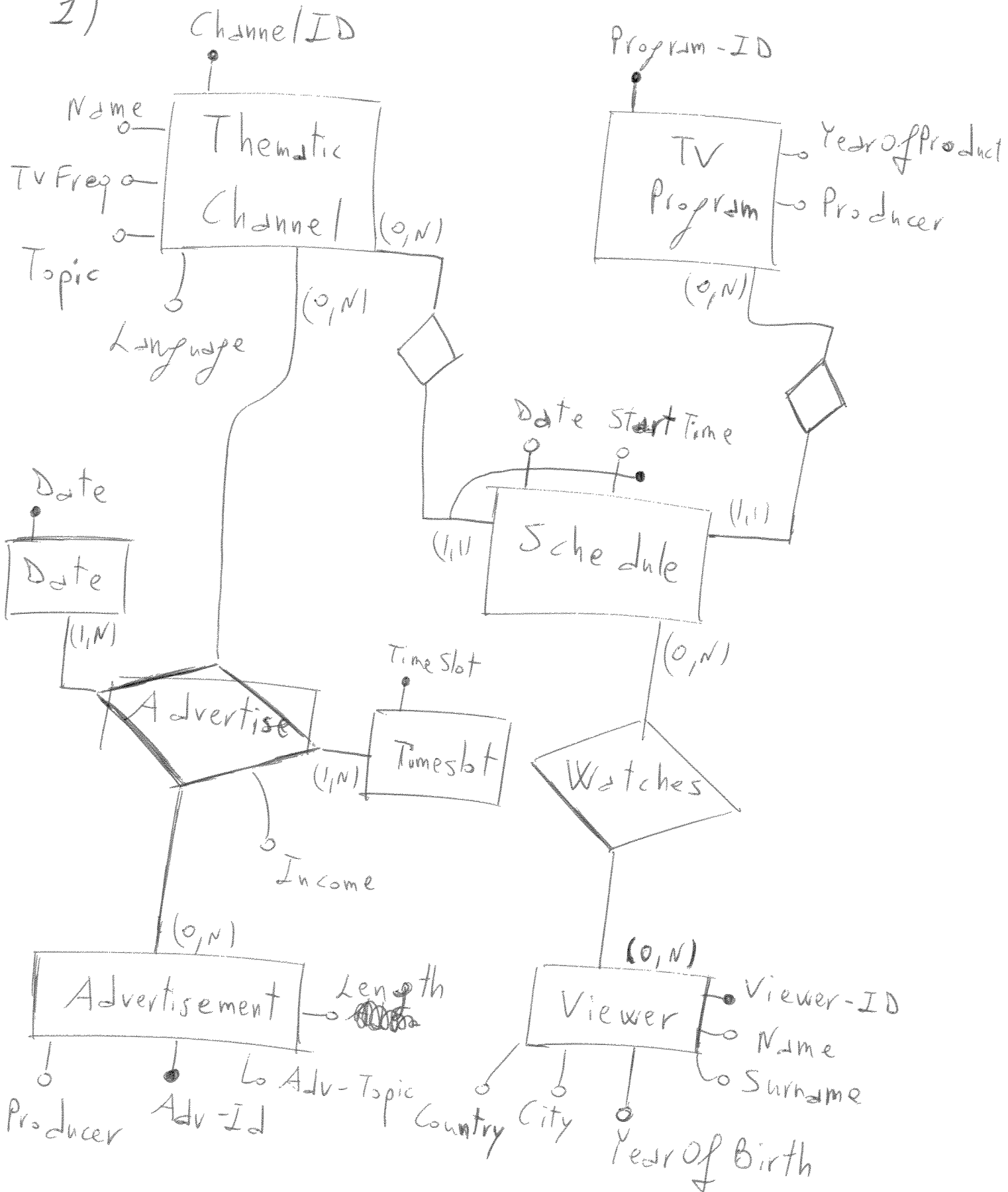Duration: 1h45min

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)

Thematic Channel
- ChannelID
- Name
- TvFreq
- Topic
- Language

(O,N)

(O,N)

TV Program
- Program-ID
- YearOfProduct
- Producer

(O,N)

Date
- Date

(I,N)

Advertise
- Income

Timeslot
- Time Slot

(I,N)

Schedule
- Date
- Start Time

(I,I)

(I,I)

(O,N)

Watches

(O,N)

Advertisement
- Producer
- Adv-Id
- Adv-Topic
- Length

(O,N)

Viewer
- Viewer-ID
- Name
- Surname
- Country
- City
- Year Of Birth

(O,N)

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
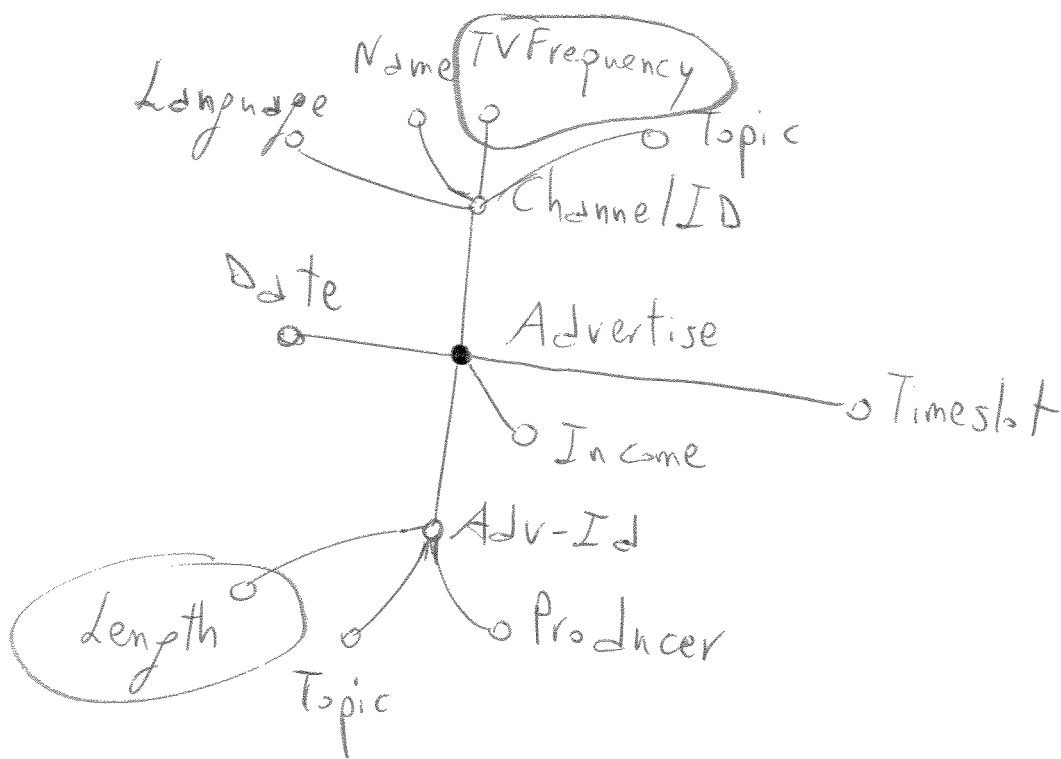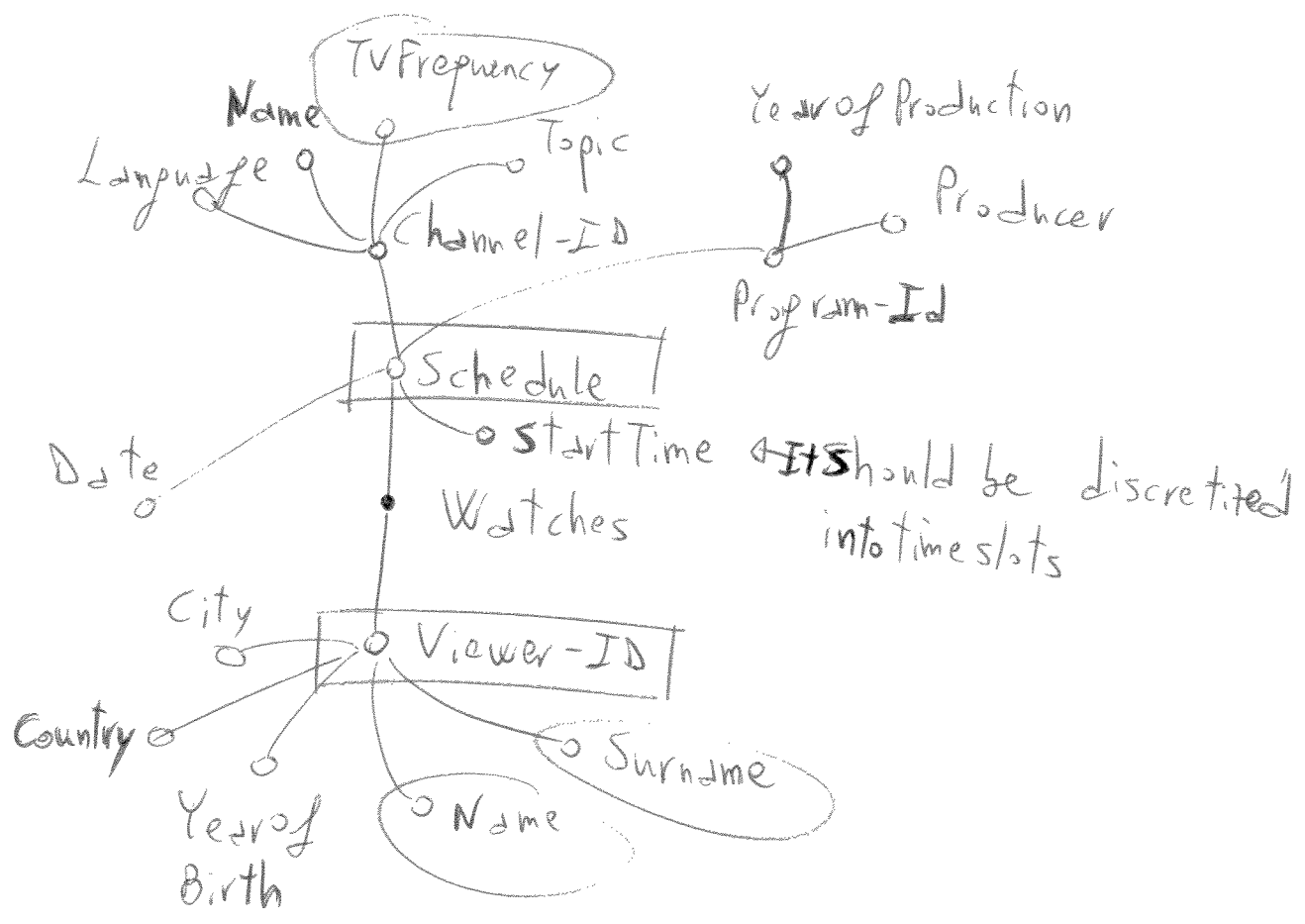
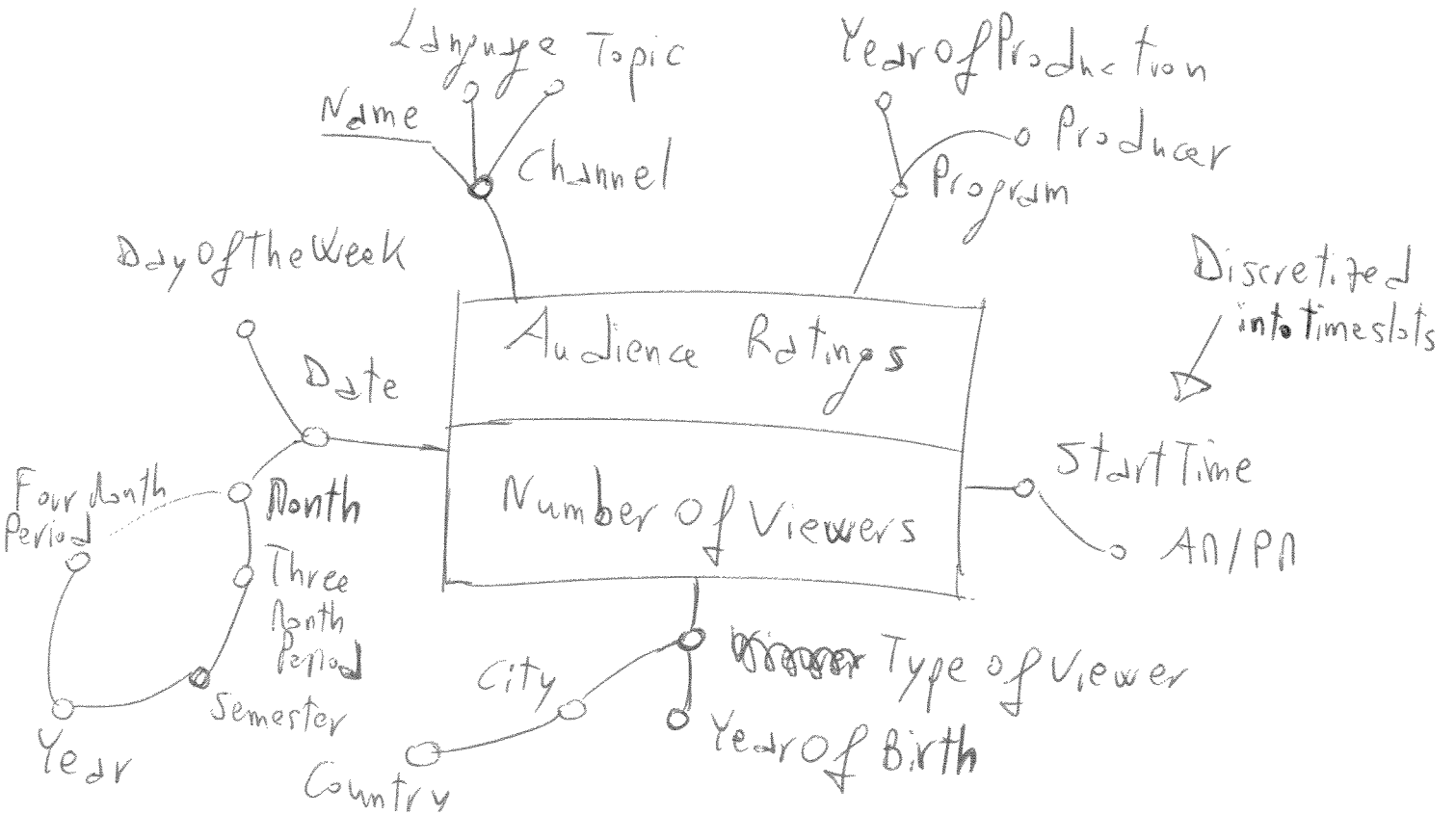# Attribute trees

Advertising

◯ = pruning

▭ = grafting

Language
Name
TVFrequency
Topic
ChannelID
Date
Advertise
Timeslot
Income
Adv-Id
Length
Producer
Topic

④

Audience Ratings

◯ = pruning

▭ = grafting

TVFrequency

Name

Language

Topic

Year of Production

Channel-ID

Producer

Program-Id

Schedule

startTime ◄ It should be discretized into time slots

Date

Watches

City

Viewer-ID

Country

Surname

Year of Birth

Name

Fact schema / Fact Model

Language Topic

Name

Channel

Advertising

Income

Number Of Advertisements

Day of the Week

Date

Month

Four Month Period

Three Month Period

Semester

Year

Topic

Advertisement

Producer

Timeslot

AN/PN

Language Topic

Name

Channel

Year of Production

Producer

Program

Audience Ratings

Number Of Viewers

Day of the Week

Date

Month

Four Month Period

Three Month Period

Semester

Year

Discretized into timeslots

Start Time

AN/PN

City

Country

Type of Viewer

Year of Birth

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.

Fact Advertising (ChannelID, DateID, AdvID, TimeID, Income, NumOfAdvs)

DimChannel (ChannelID, Name, Language, Topic)

DimDate (DateID, Day Of Week, Month, Three Month Period, Four Month Period, Semester, Year)

DimAdvertisement (AdvID, Topic, Producer)

DimeTime (TimeID, Timeslot, AM_PM)

Fact Audience (ChannelID, ProgramID, TimeID, DateID, ViewerTypeID, NumOfViewers)

Dim Program (<u>Program ID</u> , Producer, Year Of Production)

Dim Type Viewer (<u>Viewer Type ID</u> , Year Of Birth, City, Country)

2.c) <u>Sql queries</u>

(i) Select sum(Income), Day Of Week
From FactAdvertising 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 FactAdvertising 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;
        SelectDC.ChannelID, DC.Name, ProgramID
        From ChannelProgramsAudience CPA1, Dim Channel DC
        where CPA1.ChannelID = DC.ChannelID AND
        TotalAudience = (Select MAX(TotalAudience)

From Channel Programs Audience CPA2

where CPA2.ProgramID =

CPA1.ProgramID)

➡ Another possible solution (the former one is more
efficient)
Create view Channel Programs Audience (--)

-- `

=

Select DC.ChannelID, DC.Name, ProgramID

From Channel Programs Audience CPA1, Dim Channel DC

Where CPA1.ChannelID = DC.ChannelID

and CPA1.ChannelID, ~~CPA1.Program~~ Total Audience

IN (Select ChannelID, ~~Program~~ MAX(Total Audienc

From Channel Programs Audience

Group by ChannelID)