

Given Mediated Schemas

- Movie (title, director, date, genre),
- Director (director, nationality, birth_date, death_date)
- Poster (title, URL)

1. Write Local-As-View (LAV) rules that describe each source.

- i) S1 (title, genre) <--- Movie (title, director, date, genre), date < 2000
Director (director, 'America', birth_date, death_date)
- ii) S2 (title, date) <--- Movie (title, 'David Fincher', date, genre), date > 2001
Director (director, 'America', birth_date, death_date)
- iii) S3 (title, director) <--- Movie (title, director, date, 'Sci Fi'),
Director (director, nationality, birth_date, death_date)
- iv) S4 (title, URL) <--- Movie (title, director, date, genre), Poster (title, URL)

2) Given the query that searches for all the Drama movies by American directors released before 2000 that have poster URLs online. The returned results should be the titles of the movies and poster URLs. Write the query using the mediated schema and reformulate the query in LAV using the Bucket algorithm. Show the derivations for each step.

Q1(title, URL) ←
Movie (title, director, date, 'Drama'), Director (director, 'America', birth_date, death_date),
Poster (title, URL), date < 2000

Step 1: Filling the Buckets :

Sources for Movie predicate : S1(title, genre), S4(title, URL)
S3 can not be included because of the clash with the genre.
S2 can not be included because of the clash with the dates.

Sources for Director predicate : S1(title, genre), S2(title, date), S3(title, director)
Sources for Poster predicate : S4(title, URL)

Check all the combinations for containment:

Combination i: with substitution genre {genre | 'drama'}
{S1(title, genre), S1(title, genre')} S4(title, URL) <---
Movie (title, director, date, 'drama'), date < 2000
Director (director, 'America', birth_date, death_date)
Movie (title, director, date, 'drama'), date < 2000
Director (director, 'America', birth_date, death_date)
Movie (title, director, date, 'drama'), Poster (title, URL) } ⊆ Q1

Combination ii:

```
{S1(title,genre), S2(title',date'), S4(title'', URL)}<----  
  Director (director, 'America', birth_date, death_date)  
  Movie (title, director, date, genre),date<2000  
  Movie (title', 'David Fincher', date', genre'), date'>2001  
                                     Director (director', 'America', birth_date', death_date')  
Poster (title'', URL)   }
```

There is a clash between date and date', so this combination can be included.

Combination iii:

```
{S1(title,genre), S3(title',director'), S4(title'', URL)}<-----  
{  
  Movie (title, director, date, genre), date<2000  
                                     Director (director, 'America', birth_date, death_date)  
  Movie (title,director', date', 'Sci Fi'),  
                                     Director (director', nationality', birth_date', death_date')  
  Movie (title, director'', date'', genre'),Poster (title, URL)  
}
```

Containment test fails with respect to genre, so this combination can not be included.

Combination iv:

```
{S4(title,URL),S1(title,genre),S4(title, URL) with the substitution {genre|'Drama'}}<---  
  Movie (title, director, date, 'Drama'),Poster (title, URL)  
  Movie (title, director', date', 'Drama'), date<2000  
                                     Director (director', 'America', birth_date, death_date)  
  Movie (title, director'', date'', 'Drama'),Poster (title, URL')   }  $\subseteq$  Q1
```

Combination v :

```
{S4(title,URL),S2(title,date),S4(title, URL)}<---  
  
S4(title,URL),S1(title,genre),S4(title, URL),  
Movie (title,'David Fincher', date, genre), date>2001  
                                     Director (director, 'America', birth_date, death_date)  
S4(title,URL),S1(title,genre),S4(title, URL)}
```

The containment check fails because of date, so this combination can not be included.

Combination vi:

```
S4(title, URL), S3(title, director), S4(title, URL) <---
{
  Movie (title, director, date, genre), Poster (title, URL),
  Movie (title, director, date, 'Sci Fi'),
      Director (director, nationality, birth_date, death_date)
  Movie (title, director, date, genre), Poster (title, URL)
}
```

The containment test fails because of genre.

From the union of above valid combinations, the sources that give maximum containment
{S1(title, 'Drama'), s4(title, URL)}

3. For the same query as in question 2 give the inverse rules program that answers the query, and simplify the program

```
Q1(title, URL) ←
  Movie (title, director, date, 'Drama'), Director (director, 'America', birth_date, death_date),
  Poster (title, URL), date < 2000
```

Local Mappings :

- i) $S1(\text{title}, \text{genre}) \subseteq \text{Movie}(\text{title}, \text{director}, \text{date}, \text{genre}), \text{date} < 2000$
Director (director, 'America', birth_date, death_date)
- ii) $S2(\text{title}, \text{date}) \subseteq \text{Movie}(\text{title}, \text{'David Fincher'}, \text{date}, \text{genre}), \text{date} > 2001$
Director (director, 'America', birth_date, death_date)
- iii) $S3(\text{title}, \text{director}) \subseteq \text{Movie}(\text{title}, \text{director}, \text{date}, \text{'Sci Fi'}),$
Director (director, nationality, birth_date, death_date)
- iv) $S4(\text{title}, \text{URL}) \subseteq \text{Movie}(\text{title}, \text{director}, \text{date}, \text{genre}), \text{Poster}(\text{title}, \text{URL})$

i) $\forall \text{title}, \forall \text{genre} [S1(\text{title}, \text{genre}) \Rightarrow \exists \text{director}, \exists \text{date}, \exists \text{birth_date}, \exists \text{death_date} (\text{Movie}(\text{title}, \text{director}, \text{date}, \text{genre}), \text{date} < 2000$
Director (director, 'America', birth_date, death_date))]

Substitutions :

```
{genre|'Drama'}
{nationality|'America'}
```

Skolemization functions :

- director = f1(title, genre)
- date = g1(title, genre)

- `birth_date = h1(title, genre)`
- `death_date = i1(title, genre)`

Inverse rules :

- INV1 : s1(title, 'Drama') -> Movie(title, f1(title,'Drama'), g1(title, 'Drama'), 'Drama'),
- INV2 : s1(title, 'Drama') -> Director (f1(title,'Drama'), 'America',h1(title, 'Drama'), i1(title, genre))
- INV3 : s1(title, 'Drama') -><(h1(title, 'Drama'),2000)

ii) $\forall \text{title}, \forall \text{genre} [\text{S1}(\text{title}, \text{date}) \Rightarrow \exists \text{director}, \exists \text{genre}, \exists \text{birth_date}, \exists \text{death_date} (\text{Movie}(\text{title}, \text{'David Fincher'}, \text{date}, \text{genre}), \text{date} > 2001 \wedge \text{Director}(\text{director}, \text{'America'}, \text{birth_date}, \text{death_date}))]$

This can not be used because of date clash.

iii) $\forall \text{title}, \forall \text{director} [\text{S1}(\text{title}, \text{director}) \Rightarrow \exists \text{date}, \exists \text{nationality}, \exists \text{birth_date}, \exists \text{death_date} (\text{Movie}(\text{title}, \text{director}, \text{date}, \text{'Sci Fi'}, \text{date} > 2001$
Director (\text{director}, \text{nationality}, \text{birth_date}, \text{death_date}))]

This can not be used because of genre clash.

iv) $\forall \text{title}, \forall \text{URL} [\text{S4}(\text{title}, \text{URL}) \subseteq \exists \text{date}, \exists \text{genre}, \exists \text{director} (\text{Movie}(\text{title}, \text{director}, \text{date}, \text{genre}), \text{Poster}(\text{title}, \text{URL}))]$

Skolemization functions :

- genre = f4(title,URL)
- date = g4(title, URL)
- director = h4(title, URL)

Inverse rules :

- INV4 : s4(title, URL) -> Movie (title, h4(title, URL), g4(title, URL), f4(title, URL))
- INV5 : s4(title, URL) -> Poster (title, URL)

Q:-Movie (title, director, date, 'Drama'), Director (director, 'America', birth_date, death_date), Poster (title, URL), date<2000

Simplification :

Q:-s1(title, 'Drama'), s1(title, 'Drama'),s4(title, URL) ,s1(title, 'Drama')

Q:-s1(title, 'Drama'),s4(title, URL)