Databases Project – Spring 2017

Prof. Anastasia Ailamaki

Team No:

Names: Romain Gehrig, Dominique Roduit, Thierry Treyer

Contents

Contents 1

Deliverable 1 2

Assumptions 2

Entity Relationship Schema 2

Schema 2

Description 2

Relational Schema 2

ER schema to Relational schema 2

DDL 3

General Comments 3

Deliverable 2 4

Assumptions 4

Data Loading 4

Query Implementation 4

Query a: 4

Description of logic: 4

SQL statement 4

Interface 4

Design logic Description 4

Screenshots 4

General Comments 4

Deliverable 3 5

Assumptions 5

Query Implementation 5

Query a: 5

Description of logic: 5

SQL statement 5

Query Analysis 5

Selected Queries (and why) 5

Query 1 5

Query 2 5

Query 3 5

Interface 6

General Comments 6

# Deliverable 1

## Assumptions

<In this section write down the assumptions you made about the data. Write a sentence for each assumption you made>

**Story**

We exploded the attributes of this table in many tables, especially for all the fields containing multiples comma separated values. We describe above the replacement of attributes by new tables.

**script, pencils, inks, colors, letters**

These 5 attributes were replaced by 2 new tables :

1. **person**

this table will contain all the persons (artists or editors) who were contained in the 5 attributes above (and in the deleted attribute “editing”, explanation below). It is thus made of the attributes “alias\_id” who refers to a person, to reproduce the fact that sometimes some artists have an alias name. “firstname” and “lastname”.

1. **Participate**

This is the link between person and story to attribute artists to the stories. This single table fully replace the 5 attributes above, each different attribute can be identified by the “role” field whose the type is an enum which express which of the 5 attribute is concerned by the record.

**characters, feature**

2 attributes replaces by 3 tables. On table “hero” which will store the list of all heros, and one intermediate table by attribute which allow to split each comma separated values in many single record linked to the story.

**genre**

This attribute was replaced by two tables : one table “genre” to contain all the genre and avoid duplicata, and an intermediate table “story\_genre” with two identifying relationships, in order to explode the separated values in multiples records. Each genre associated to a story is a record in this table.

**editing**

this attribute was deleted because we can retrieve it via the “editing\_id” attribute linked to its indicia\_publisher, itself link to the issue containing the story. The editor is now in a separated table “edit” containing only editors and linking a person (an editor in this case) to the issue containing the story published by the indicia\_publisher linked to the concerned story. The credits are stored in the “note” attribute of this table.

**Issue**

**valid\_isbn**

this attribute was removed, we only kept “isbn”.

**price**

attribute replaced by a new table “price” containing the issue concerned by the price, the amount (price itself) and the currency in which this price is exprimed (currency is simply an enum, since the values of currency are a finite set of same values).

## Entity Relationship Schema

<In this section you should have figure of the ER schema as well as descriptions about entities and relations>

### Schema



### Description <Describe all the choices you made for Entities and Relationships>

**Story (relations)**

**story.id - story\_reprint.origin\_id**

An origin story (old printed story) is 0 or one story.

A story can be the origin story 0, 1 or many times (if reprinted many times).

Identifying relationship because the origin story printing is itself a story.

**story.id - story\_reprint.target\_id**

A target story (old printed story) is exactly one story.

A story can be the target story 1 or many times (if reprinted many times).

Identifying relationship because the target story printing is itself a story.

**story.type\_id - story\_type.id**

A story can be of zero or one type.

A story type could be attributed to 0 or many stories.

**story.issue\_id – issue.id**

A story is contained by 1 issue

An issue contains 1 or many issues.

**story.id – participate.story\_id**

A person can participate in one or many stories

A story can have one or many artists who participated to its realisation

**story.id – feature.story\_id**

A feature can be part of zero or one story

A story can have none or many feature.

**story.id – character.story\_id**

A character can be attributed to none or many stories

A story have at least one characters, or many.

**story.id – story\_genre.story\_id**

A genre can be attributed to 1 or many stories

A story can be of one or many genre

**Indicia\_publisher.id – issue.indicia\_publisher\_id**

A story can be published by 0 or 1 indicia publisher

An indicia publisher can publish 1 or many stories

**Issue (relations)**

**Issue\_reprint.origin\_id – issue.id**

An origin issue (old printed issue) is exactly one issue.

An issue can be the origin issue 1 or many times (if reprinted many times).

Identifying relationship because the origin issue is itself an issue.

**Issue\_reprint.target\_id – issue.id**

A target (new printed issue) is exactly one issue.

An issue can be the target issue 1 or many times (if reprinted many times).

Identifying relationship because the target issue is itself an issue.

**Indicia\_publisher.id – issue.indicia\_publisher\_id**

An indicia\_publisher can publish 1 or many issue.

An issue can be published by 0 or one indicia publisher

**issue.id – price.issue\_id**

An issue can be sold to 1 or many price

An a price with its currency can be attributed to only one issue.

**publisher (relations)**

**Publisher.id – indicia\_publisher.publisher\_id**

An indicia publisher can be owned by only one publisher.

A publisher can be the owner of 0 or many indicia publishers.

Identifying relationship because an indicia\_publisher depends directly of a publisher, it can't exists without a publisher.

**Publisher.id – brand\_group.publisher\_id**

A brand is owned by exactly one publisher.

A publisher can own 0 or many brands.

Identifying relationship because an brand\_group depends directly of a publisher, it can't exists without a publisher.

**Publisher.id – series.publisher\_id**

A publisher can publish 0, one or many series.

A serie can be published by one publisher.

**Country (relations)**

**Indicia\_publisher.country\_id – country.id**

An indicia publisher comes from one country

A country can be attributed to 0 or many company

**Country.id – publisher.country\_id**

A publisher comes from one country.

A country can be assign to many publisher.

**Series.country\_id – country.id**

A country can be attributed to 0, one or many series.

A serie comes from one country.

**Language (relations)**

**Language.id – series.language\_id**

A language can be attributed to 0, one or many series.

A series can be in one language.

**Series (relations)**

**Series\_publication\_type.id – series.publication\_type\_id**

A publication type can be attributed to 0, one or many series.

A serie can be of one publication type.

**Series.id – issue.series\_id**

An issue belongs to 0 or one serie.

A serie can wrap 1 or many issues.

**Issue.id – series.first\_issue\_id**

It could exists 0 or 1 first issue for a serie

An issue can be the first issue of 0, 1 or many series.

It's weird but according the data in the csv file, first and last issue both can be NULL.

**Issue.id – series.last\_issue\_id**

It could exists 0 or 1 last issue for a serie

An issue can be the last issue of 0, 1 or many series.

It's weird but according the data in the csv file, first and last issue both can be NULL.

## Relational Schema

### ER schema to Relational schema

<Describe the transition from ER schema to Relational schema>



The relations made only of single strokes as above (thin or bold but without arrows), generate the creation of a new table containing two foreign keys who link together the two tables concerned.

C:\Users\Dominique\AppData\Local\Microsoft\Windows\INetCache\Content.Word\notnullrelation.png

The relations who contains one arrow are reduced to a NON NULL foreign key in the table in front of the arrow. In the example above, the relation could be translated by a NON NULL foreign key in the table B.



In the relations who are only made of one-one (two bold arrows) and zero-one (one bold and one thin arrows as in this example) constraints, the foreign key must be UNIQUE.

In the example above, the relation could be translated by a UNIQUE foreign key in the table A.

The participation constraints “at least one” (denoted by a bold stroke) can’t be captured in SQL without check constraint or assertion.

### 

### DDL

<Provide the DDL>

## General Comments

<In this section write general comments about your deliverable (comments and work allocation between team members>

# Deliverable 2

## Assumptions

<In this section write down the assumptions you made about the data. Write a sentence for each assumption you made>

## Data Loading

## Query Implementation

<For each query>

### Query a:

#### Description of logic:

<What does the query do and how do I decide to solve it>

#### SQL statement

<The SQL statement>

## Interface

### Design logic Description

<Describe the general logic of your design as well as the technology you decided to use>

### Screenshots

<Provide some initial screen shots of your interface>

## General Comments

<In this section write general comments about your deliverable (comments and work allocation between team members>

# Deliverable 3

# Assumptions

<In this section write down the assumptions you made about the data. Write a sentence for each assumption you made>

## Query Implementation

<For each query>

### Query a:

#### Description of logic:

<What does the query do and how do I decide to solve it>

#### SQL statement

<The SQL statement>

## Query Analysis

### Selected Queries (and why)

#### Query 1

<Initial Running time:

Optimized Running time:

Explain the improvement:

Initial plan

Improved plan>

#### Query 2

<Initial Running time:

Optimized Running time:

Explain the improvement:

Initial plan

Improved plan>

#### Query 3

<Initial Running time:

Optimized Running time:

Explain the improvement:

Initial plan

Improved plan>

# Interface

### Design logic Description

<Describe the general logic of your design as well as the technology you decided to use>

### Screenshots

<Provide some initial screen shots of your interface>

# General Comments

<In this section write general comments about your deliverable (comments and work allocation between team members>