Video Game Database

1- Identify and explain your database

This database stores information about the top 11,000 or so video games that have had the most sales of all time. It captures video game sales in North America, Europe, Japan, and other regions grouped as "Other". It has games from various platforms: PC, recent consoles, and old consoles.

The sample data was obtained by a free online data source. There are about 11,000 games, 31 platforms, 577 publishers, and 4 regions in this dataset.

What do all of these tables and columns mean?

Table Explanations

Genre

This table has a short list of game genres, such as Strategy or Racing. It's helpful if you want to see what genre a game is, or the types of genres that are making sales.

Game

A list of all of the games in the database. It only contains a title, and a genre_id to link to the genre table. There is no concept of a series in this data (e.g. "Call of Duty" or "Need for Speed").

Publisher

A company that publishes video games.

Game_Publisher

This is a joining table that defines the games and their publishers, because in the source data, a game can have many publishers. One example is different publishers work with the game on different platforms (e.g. one on PC and one on XBOX).

Platform

A list of the different gaming platforms in the source data. For example, PC, Playstation 3, XBOX360, Nintendo.

Game_Platform

This table links the games, publishers, and the platform. Using this table, you can see the games, who published them, and on which platform (when you join to the related tables).

It also includes the release_year. Why is the release_year here and not on the game table?

This is because, in the source data, a game can be released by different publishers on a different platform in a different year. For example, SimCity 2000 was released on PC in 1992 and PS in 1996.

Region

This lists 4 different regions that sales were captured in: North America, Europe, Japan, and Others.

Region_Sales

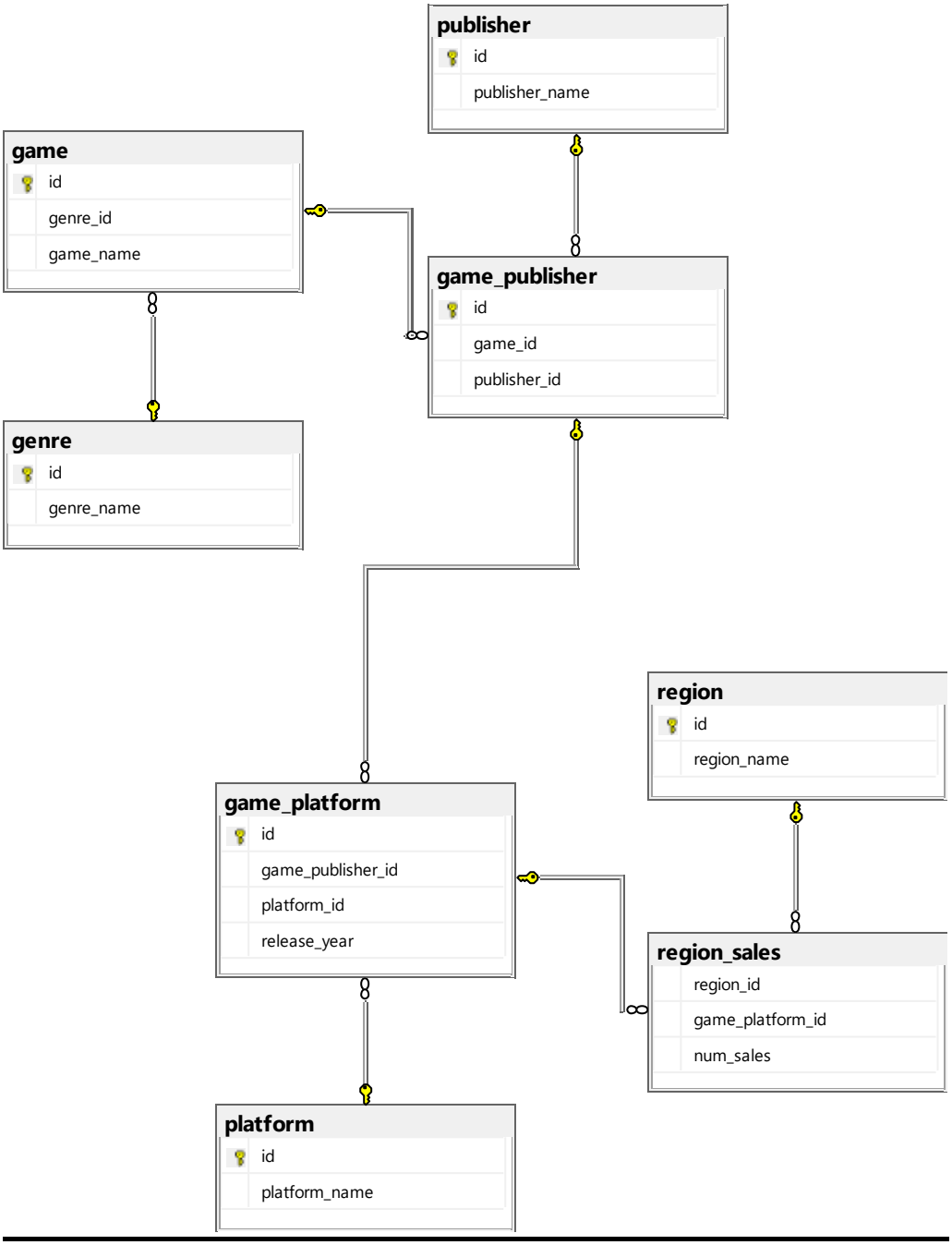
This table shows all sales made for games in each region. It's the main table used to calculate the sales for queries.

The num_sales column is the number of sales of the game in the region, and **the number is in millions**.

2- What are the measures you have in the database

- Region_Sales is the fact table
- It has the measures (num of sales)

3- Diagram



4- What KPIs are you going to measure ?

- Total Sales
- Sales by Region
- Sales by Genre
- Sales by Platform
- Publisher Market Share
- Top Selling Games
- Sales Growth across time
- Average Sales per Game
- Sales Concentration
- Platforms that use different types of genres
- The number of games produced for each genre

5- Write SQL query that will be used to achieve the KPIs

1- Total Sales :

```
SELECT SUM(num_sales) AS Total_Sales
FROM Region_Sales;
```

2- Sales by Region :

```
SELECT region_name, SUM(num_sales) AS Sales
FROM Region_Sales rs
LEFT JOIN Region r
ON r.id = rs.region_id
GROUP BY region_name
ORDER BY 2 DESC;
```

3- Sales by Genre :

```
SELECT Genre.genre_name, SUM(Region_Sales.num_sales) AS Sales
FROM Game
JOIN Game_Publisher ON game_publisher.game_id = Game.id
JOIN Game_Platform ON Game_Platform.game_publisher_id = Game_Publisher.id
JOIN Genre ON Genre.id = Game.genre_id
JOIN Region_Sales ON Region_Sales.game_platform_id = Game_Platform.id
GROUP BY Genre.genre_name
ORDER BY 2 DESC;
```

4- Sales by Platform :

```
SELECT Platform.platform_name, SUM(Region_Sales.num_sales) AS Sales
FROM Platform
JOIN Game_Platform ON Game_Platform.platform_id = Platform.id
JOIN Region_Sales ON Region_Sales.game_platform_id = Game_Platform.id
GROUP BY Platform.platform_name
ORDER BY 2 DESC;
```

5- Publisher Market Share :

```
SELECT Publisher.publisher_name, SUM(Region_Sales.num_sales) AS Sales,
SUM(Region_Sales.num_sales) / (SELECT SUM(num_sales) FROM Region_Sales) AS Market_Share
FROM Game_Publisher
JOIN Publisher ON Publisher.id = Game_Publisher.publisher_id
JOIN game_platform ON game_platform.game_publisher_id = Game_Publisher.id
JOIN Region_Sales ON Region_Sales.game_platform_id = game_platform.id
GROUP BY Publisher.publisher_name
ORDER BY 2 DESC;
```

6- Top Selling Games :

```
SELECT Game.game_name, SUM(Region_Sales.num_sales) AS Sales
FROM Game
JOIN Game_Publisher ON game_publisher.game_id = Game.id
JOIN Game_Platform ON Game_Platform.game_publisher_id = Game_Publisher.id
JOIN Region_Sales ON Region_Sales.game_platform_id = Game_Platform.id
GROUP BY Game.game_name
ORDER BY SUM(Region_Sales.num_sales) DESC;
```

7- Sales Growth across years :

```
SELECT Game_Platform.release_year, SUM(Region_Sales.num_sales) AS Sales
FROM Game_Platform
JOIN Region_Sales ON Region_Sales.game_platform_id = Game_Platform.id
GROUP BY Game_Platform.release_year
ORDER BY 2 DESC;
```

8- Average Sales per Game :

```
SELECT AVG(num_sales) AS Average_Sales
FROM Region_Sales;
```

9- Sales Concentration :

```
SELECT SUM(num_sales) / (SELECT SUM(num_sales) FROM Region_Sales) AS
Sales_Concentration
FROM (
   SELECT Top (10) num_sales
   FROM Region_Sales
   ORDER BY num_sales DESC

) AS Top_Selling_Games;
```

10 - Platforms that use different types of genres :

```
SELECT DISTINCT platform.platform_name,COUNT(DISTINCT genre.genre_name) AS
Num_of_Genres
FROM platform
JOIN game_platform ON game_platform.platform_id = platform.id
JOIN game_publisher ON game_publisher.id = game_platform.game_publisher_id
JOIN game ON game.id = game_publisher.game_id
JOIN genre ON genre.id = game.genre_id
GROUP BY platform.platform_name
HAVING COUNT(DISTINCT genre.genre_name) > 1
ORDER BY 2 DESC;
```

11- The number of games produced for each genre :

```
SELECT genre.genre_name, COUNT(game.id) AS num_games
FROM genre
JOIN game ON game.genre_id = genre.id
GROUP BY genre.genre_name
ORDER BY 2 DESC;
```