

Data Lake Project

Data lake of Youtube Comments

Student: Mohamed Sabkhi
Class: 2IDSD1

Dataset source

The Dataset was scraped using the youtube API.

The dataset was uploaded to kaggle by NIPUN ARORA.

Link of the dataset: [most liked comments on youtube](#)

Youtube Comments Dataset preview:

	A	B	C	D	E	F	G	H
		Video Name	Channel Name	Comment Id	User Name	Comment	Likes	
1	0	Luis Fonsi - Despacito	LuisFonsiVEVO	UgwV0tapZzaFxdYm1z14AaABAg	GAMIT yuvraj	The people who liked this co	98280	
2	1	Luis Fonsi - Despacito	LuisFonsiVEVO	UgyiccbIP3A16aixpit4AaABAg	MrCrazy	- Wait, it's 7B views- Always	773	
3	2	Luis Fonsi - Despacito	LuisFonsiVEVO	Ugx7THW3EBNJIhJuwSB4AaABAg	60K Subs With 1 Video Cha	*Teacher: What is the popul	356081	
4	3	Luis Fonsi - Despacito	LuisFonsiVEVO	UgzXUsl6yrRjTKNASf94AaABAg	Srivathsa D	Let's be honest this wasn't y	13	
5	4	Luis Fonsi - Despacito	LuisFonsiVEVO	UgxCu6mywg6uqjueHQ14AaABAg	1k Subscribers Challenge V	Types Of People:10% Enjoyir	122	
6	5	Luis Fonsi - Despacito	LuisFonsiVEVO	UgzryH8U6Dz_yBmlgEZ4AaABAg	LR ICED WOLF	3.2 Million comments if you	370547	
7	6	Luis Fonsi - Despacito	LuisFonsiVEVO	UgzPg7VsuRTwJf77jch4AaABAg	lucas tehe12	claim your "here before 7 bil	763	
8	7	Luis Fonsi - Despacito	LuisFonsiVEVO	Ugw61yKNdyVJ5T4R_sN4AaABAg	MemeDille	The ones who are NOT from	94	
9	8	Luis Fonsi - Despacito	LuisFonsiVEVO	UgzhucoG3QOV4AbAYAt4AaABAg	Guty Castro soda	Song: spanishComments: En	150	
10	9	Luis Fonsi - Despacito	LuisFonsiVEVO	UgxaUPmMyW0KD8kqi4B4AaABAg	Taha'nin Mutfağı	Kimler burda 🤔🤔🤔	45	
11	10	Luis Fonsi - Despacito	LuisFonsiVEVO	UgxcqxyEw2jU7KvhofV4AaABAg	cruisendude	The first viewer must feel like	36446	
12	11	Luis Fonsi - Despacito	LuisFonsiVEVO	Ugy-CwNxTeLKUuYOXdF4AaABAg	Pls HeLP Me reach 100 SU	Just imagine how many peop	142	
13	12	Luis Fonsi - Despacito	LuisFonsiVEVO	Ugz6XednoX2ElrbsRTB4AaABAg	Lokesh Prasad m.s	hello everyone 🤔 Who cam	5	
14	13	Luis Fonsi - Despacito	LuisFonsiVEVO	UgwxXGcEfq1XTFgIHdZ4AaABAg	Mikhil Das	3.5 M Comments if you find	10	
15	14	Luis Fonsi - Despacito	LuisFonsiVEVO	UgzCmYFAMK8fy5_cHWx4AaABAg	Please Help me 1K Subscri	*Let`s Go 7 Billion views > H	109	
16	15	Luis Fonsi - Despacito	LuisFonsiVEVO	UgwOBGutUxio5Z2MWMd4AaABAg	Anshu Soren	Who just came to check the	321690	
17	16	Luis Fonsi - Despacito	LuisFonsiVEVO	UgysG0YRk2F8mNTpIJh4AaABAg	lol	I'll see you guys when the "V	8	
18	17	Luis Fonsi - Despacito	LuisFonsiVEVO	UgwVH12PzPlz0EEfYVMAAaABAg	Help Me 5k Subscribers Wi	*Despacito **Youtube: This	20	

Youtube Comments dataset Columns:

Video Name: Name of the video from which the comment was extracted

Channel Name: Name of the channel from which the video was uploaded

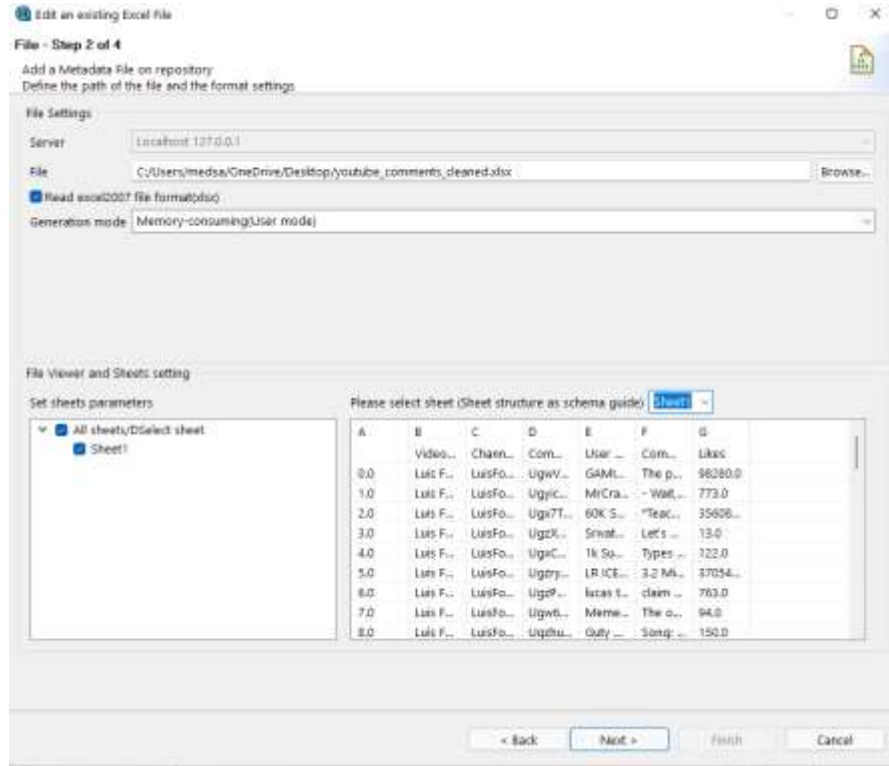
CommentID: unique identifier of the comment.

User name: user name of the commenter.

Comment: the text comment

Likes: number of likes on the comment.

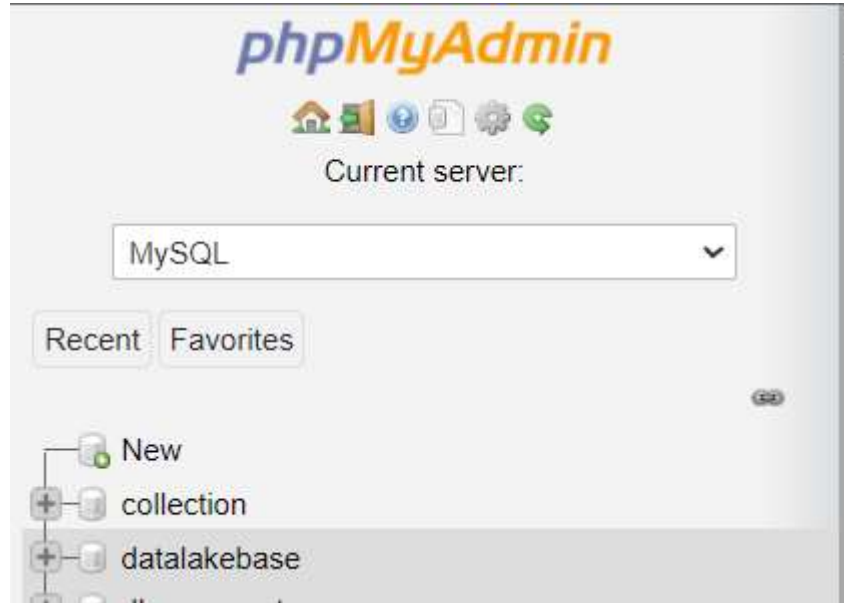
Loading Excel sheet into talend



- videoname 0.1
 - metadata
 - Columns(7)
 - Channel_Name
 - Column0
 - Comment
 - Comment_Id
 - Likes
 - User_Name
 - Video_Name

Creating MySQL database using phpmyadmin

Databasename: datalakebase



Creating tables in mysql database.

Table	Action	Rows	Type	Collation	Size	Overhead
<input type="checkbox"/> channel name	Browse Structure Search Insert Empty Drop	38,568	MyISAM	latin1_swedish_ci	1.7 MiB	-
<input type="checkbox"/> commenters	Browse Structure Search Insert Empty Drop	19,284	MyISAM	latin1_swedish_ci	887.9 KiB	-
<input type="checkbox"/> comments	Browse Structure Search Insert Empty Drop	19,284	MyISAM	latin1_swedish_ci	3.0 MiB	-
<input type="checkbox"/> video name	Browse Structure Search Insert Empty Drop	19,075	MyISAM	latin1_swedish_ci	2.3 MiB	-
4 tables	Sum	96,211	MyISAM	latin1_swedish_ci	8.0 MiB	0 B

Table 1: channelname







#	Name	Type	Collation	Attributes	Null	Default	Comments	Extra	Action
<input type="checkbox"/>	1	CommentID	varchar(40)	latin1_swedish_ci	No	None			 Change  Drop  More
<input type="checkbox"/>	2	ChannelName	text	latin1_swedish_ci	No	None			 Change  Drop  More

Table 2: commenters



	#	Name	Type	Collation	Attributes	Null	Default	Comments	Extra	Action
<input type="checkbox"/>	1	CommentID	varchar(50)	latin1_swedish_ci		No	None			 Change  Drop  More
<input type="checkbox"/>	2	user	varchar(100)	latin1_swedish_ci		No	None			 Change  Drop  More

Table 3: comments








	#	Name	Type	Collation	Nullable	Default	Comments	Data	Action
<input type="checkbox"/>	1	CommentID	varchar(50)	latin1_swedish_ci	No	None			 Change  Drop ▼ More
<input type="checkbox"/>	2	Likes	int(11)		No	None			 Change  Drop ▼ More
<input type="checkbox"/>	3	Comment	text	latin1_swedish_ci	No	None			 Change  Drop ▼ More

Table 4: video name

	#	Name	Type	Collation	Attributes	Null	Default	Comments	Extra	Action
<input type="checkbox"/>	1	CommentID 	varchar(50)	latin1_swedish_ci		No	None			 Change  Drop  More
<input type="checkbox"/>	2	Video Name	longtext	latin1_swedish_ci		No	None			 Change  Drop  More

Establishing connection with Mysql database

Database Connection

Update Database Connection - Step 2/2

You must press the Check Button to check the Database Setting

DB Type: MySQL

DB Version: MySQL 5

String of Connection: jdbc:mysql://localhost:3306/datalakebase?noDatetimeStringSync=true

Login: root

Password:

Server: localhost

Port: 3306

DataBase: datalakebase

Additional parameters: noDatetimeStringSync=true

Test connection

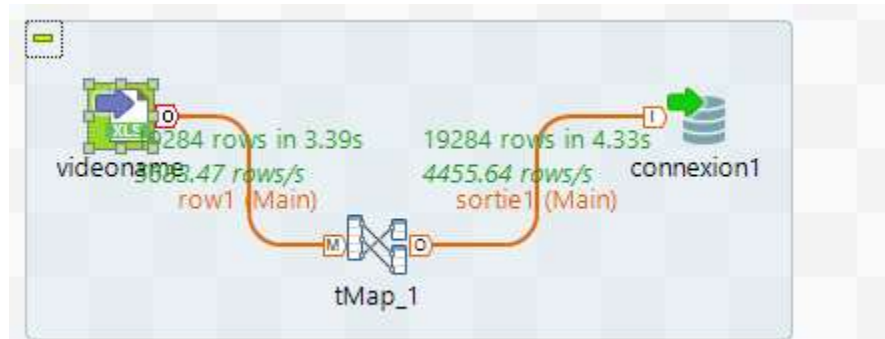
Export as context Revert Context

[How to install a driver](#)

< Back Next > Finish Cancel

Creating jobs for each table:

Job1:



row1

Column

Column0
Video_Name
Channel_Name
Comment_Id
User_Name
Comment
Likes

Var

sortie1

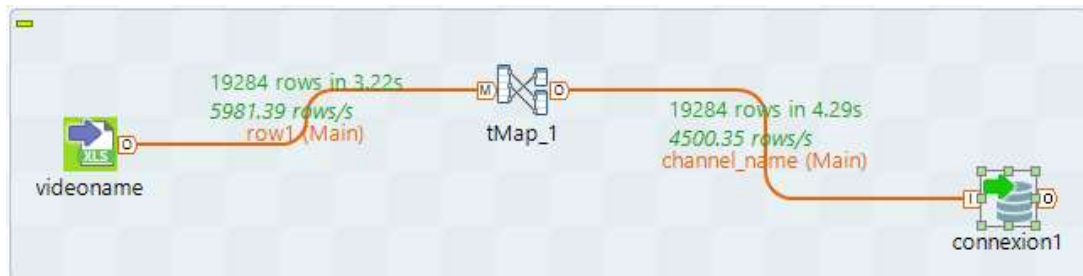
Expression

row1.Comment_Id
row1.Video_Name

Column

CommentID
Video_Name

job2:



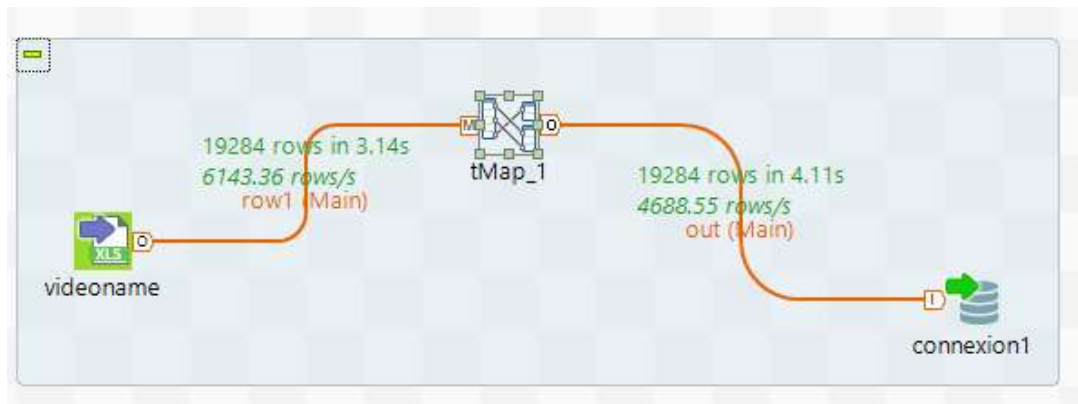
row1	
Column	
Column0	
video_Name	
Channel_Name	
Comment_Id	
User_Name	
Comment	
Likes	

Var	

vidname	
Expression	Column
	CommentID
	Video_Name

channel_name	
Expression	Column
row1.Comment_Id	CommentID
row1.Channel_Name	ChannelName

job3:

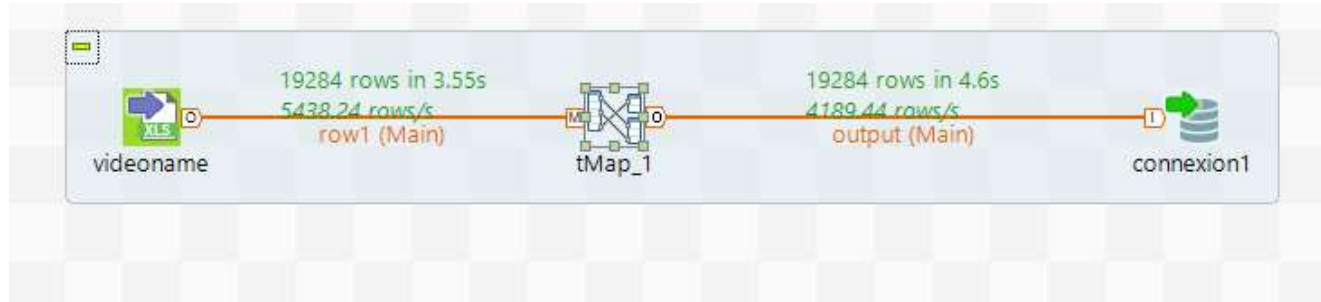


row1	
Column	
Column0	
Video_Name	
Channel_Name	
Comment_id	
User_Name	
Comment	
Likes	

Var

out	
Expression	Column
row1.Comment_id	CommentID
row1.User_Name	user

job4



row1	Var	output																
<table><tr><th>Column</th></tr><tr><td>Column0</td></tr><tr><td>Video_Name</td></tr><tr><td>Channel_Name</td></tr><tr><td>Comment_Id</td></tr><tr><td>User_Name</td></tr><tr><td>Comment</td></tr><tr><td>Likes</td></tr></table>	Column	Column0	Video_Name	Channel_Name	Comment_Id	User_Name	Comment	Likes		<table><tr><th>Expression</th><th>Column</th></tr><tr><td>row1.Comment</td><td>Comment</td></tr><tr><td>row1.Comment_id</td><td>CommentID</td></tr><tr><td>row1.Likes</td><td>Likes</td></tr></table>	Expression	Column	row1.Comment	Comment	row1.Comment_id	CommentID	row1.Likes	Likes
Column																		
Column0																		
Video_Name																		
Channel_Name																		
Comment_Id																		
User_Name																		
Comment																		
Likes																		
Expression	Column																	
row1.Comment	Comment																	
row1.Comment_id	CommentID																	
row1.Likes	Likes																	

Checking database:

Table: Channel name

SELECT * FROM `channel name`

1

>

>>

Number of rows:

25

+ Options

CommentID	ChannelName
UgwV0tapZzaFxdYm1z14AaABAg	LuisFonsiVEVO
UgyiccbIP3A16aixpit4AaABAg	LuisFonsiVEVO
Ugx7THW3EBNJlhJuwSB4AaABAg	LuisFonsiVEVO
UgzXUsl6yrRjTKNASf94AaABAg	LuisFonsiVEVO
UgxCu6mywg6uqjueHQ14AaABAg	LuisFonsiVEVO
UgzryH8U6Dz_yBmIgEZ4AaABAg	LuisFonsiVEVO
UgzPg7VsuRTwJf77jch4AaABAg	LuisFonsiVEVO
Ugw61yKNdyVJ5T4R_sN4AaABAg	LuisFonsiVEVO
UgzhueoG3QOV4AbAYAt4AaABAg	LuisFonsiVEVO
UgxaUPmMyW0KD8kqi4B4AaABAg	LuisFonsiVEVO
UgxcqxyEw2jU7KvhofV4AaABAg	LuisFonsiVEVO
Ugy-CwNxTeLKUuYOXdf4AaABAg	LuisFonsiVEVO
Ugz6XednoX2ElrbsRTB4AaABAg	LuisFonsiVEVO
UgwxXGcEfq1XTFgIHdZ4AaABAg	LuisFonsiVEVO
UgzCmYFAMK8fY5_cHWx4AaABAg	LuisFonsiVEVO
UgwOBGutUxio5Z2MWMd4AaABAg	LuisFonsiVEVO
UgysG0YRk2F8mNTplJh4AaABAg	LuisFonsiVEVO
UgwVHJ2PtPLzQFFJYM94AaABAg	LuisFonsiVEVO
UgxMNFdhkeVqZluojt4AaABAg	LuisFonsiVEVO
UgyXh-VgAqQWuK59myl4AaABAg	LuisFonsiVEVO
UgwLJL7bGoqHHb_qNh4AaABAg	LuisFonsiVEVO
Ugw2FNGc54AaABAg	LuisFonsiVEVO

Console

Table: commenters

Showing rows 0 - 24 (19284 total, Query took 0.0023 seconds.)

```
SELECT * FROM `commenters`
```

1 > >> | Number of rows: 25 | Filter rows: Search this

+ Options

CommentID	user
UgwV0tapZzaFxdYm1z14AaABAg	GAMIT yuvraj
UgyiccbIP3A16aixpit4AaABAg	MrCrazy
Ugx7THW3EBNJlhJuwSB4AaABAg	60K Subs With 1 Video Challenge
UgzXUsl6yrRjTKNASf94AaABAg	Srivathsa D
UgxCu6mywg6uqjueHQ14AaABAg	1k Subscribers Challenge Without Any Video
UgzryH8U6Dz_yBmlgEZ4AaABAg	LR ICED WOLF
UgzPg7VsuRTwJf77jch4AaABAg	lucas tehe12
Ugw61yKNdyVJ5T4R_sN4AaABAg	MemeDille
UgzhueoG3QOV4AbAYAt4AaABAg	Guty Castro soda
UgxaUPmMyW0KD8kqi4B4AaABAg	Taha'n?n Mutfa??
UgxcqxyEw2jU7KvhoFV4AaABAg	cruisendude
Ugy-CwNxTeLKUuYOXdF4AaABAg	Pls HeLP Me reach 100 SUBscribers
Ugz6XednoX2ElrbsRTB4AaABAg	Lokesh Prasad m.s
UgwxXGcEfQ1XTFglHdZ4AaABAg	Mikhail Das
UgzCmYFAMK8fY5_cHWx4AaABAg	Please Help me 1K Subscribers
UgwOBGutUxio5Z2MWMd4AaABAg	Anshu Soren
UgysG0YRk2F8mNTpIJh4AaABAg	lol
UgwVHJ2PtPLzQFFJYM94AaABAg	Help Me 5k Subscribers With Videos Challenge
UgxMNFdhkeVqZluojzt4AaABAg	the funny legend
UgyXh-VgAqQWuK59myl4AaABAg	LORD MAGA
UgwJL_ZhGoqHHb_qNh4AaABAg	Captain doorway
UgwJL_ZhGoqHHb_qNh4AaABAg	Captain doorway

Console

Table: comments

✓ Showing rows 0 - 24 (19284 total, Query took 0.0035 seconds.)

```
SELECT * FROM `comments`
```

1 ▾

> >>

Number of rows:

25 ▾

Filter rows:

Search this table

+ Options

CommentID	Likes	Comment
UgwV0tapZzaFxdYm1z14AaABAg	98280	The people who liked this comment is officially be...
UgyiccbIP3A16aixpit4AaABAg	773	- Wait, it's 7B views - Always has been
Ugx7THW3EBNJlhJuwSB4AaABAg	356081	*Teacher: What is the population of the Earth?*
UgzXUsl6yrRjTKNASf94AaABAg	13	*... Let's be honest this wasn't your recommendation yo...
UgxCu6mywg6uqjueHQ14AaABAg	122	Types Of People: 10% Enjoying Song 90% Checking V...
UgzryH8U6Dz_yBmlgEZ4AaABAg	370547	3.2 Million comments if you find mine your a legen...
UgzPg7VsuRTwJf77jch4AaABAg	763	claim your "here before 7 billion" tickets here, t...
Ugw61yKNdyVJ5T4R_sN4AaABAg	94	The ones who are NOT from Tik-Tok can like this co...
UgzhueoG3QOV4AbAYAt4AaABAg	150	Song: spanish Comments: English Hotel: travago
UgxaUPmMyW0KD8kqi4B4AaABAg	45	Kimler burda??????
UgxcqxyEw2jU7KvhofV4AaABAg	36446	The first viewer must feel like the king of the wo...
Ugy-CwNxTeLKUuYOXdF4AaABAg	142	Just imagine how many people's come here everyday....
Ugz6XednoX2ElrbsRTB4AaABAg	5	hello everyone ?? Who came here to see the likes a...
UgwxXGcEfq1XTFglHdZ4AaABAg	10	3.5 M Comments if you find mine you are a legend

Table video name:

Showing rows 0 - 24 (19075 total, Query took 0.0014 seconds.)

```
SELECT * FROM `video name`
```

1 > >> | Number of rows: 25 Filter rows: Search this table Sort by

+ Options

		CommentID	Video Name
<input type="checkbox"/>	Edit Copy Delete	Ugig2D1BVXUeNHgCoAEC	Alan Walker - Faded
<input type="checkbox"/>	Edit Copy Delete	Ultra Music	1973
<input type="checkbox"/>	Edit Copy Delete	UgwV0tapZzaFxdYm1z14AaABAg	Luis Fonsi - Despacito ft. Daddy Yankee
<input type="checkbox"/>	Edit Copy Delete	UgyiccbIP3A16aixpit4AaABAg	Luis Fonsi - Despacito ft. Daddy Yankee
<input type="checkbox"/>	Edit Copy Delete	Ugx7THW3EBNJlhJuwSB4AaABAg	Luis Fonsi - Despacito ft. Daddy Yankee
<input type="checkbox"/>	Edit Copy Delete	UgzXUsl6yrRjTKNASf94AaABAg	Luis Fonsi - Despacito ft. Daddy Yankee
<input type="checkbox"/>	Edit Copy Delete	UgxCu6mywg6uqiueHQ14AaABAg	Luis Fonsi - Despacito ft. Daddy Yankee
<input type="checkbox"/>	Edit Copy Delete	UgzryH8U6Dz_yBmIgeZ4AaABAg	Luis Fonsi - Despacito ft. Daddy Yankee
<input type="checkbox"/>	Edit Copy Delete	UgzPg7VsuRTwJf77jch4AaABAg	Luis Fonsi - Despacito ft. Daddy Yankee
<input type="checkbox"/>	Edit Copy Delete	Ugw61yKNdyVJST4R_sN4AaABAg	Luis Fonsi - Despacito ft. Daddy Yankee
<input type="checkbox"/>	Edit Copy Delete	UgzhueoG3QOV4AbAYAt4AaABAg	Luis Fonsi - Despacito ft. Daddy Yankee
<input type="checkbox"/>	Edit Copy Delete	UgxaUPmMyW0KD8kqi4B4AaABAg	Luis Fonsi - Despacito ft. Daddy Yankee
<input type="checkbox"/>	Edit Copy Delete	UgxcqxyEw2jU7KvhofV4AaABAg	Luis Fonsi - Despacito ft. Daddy Yankee
<input type="checkbox"/>	Edit Copy Delete	Ugy-CwNxTeLKUuYOXdF4AaABAg	Luis Fonsi - Despacito ft. Daddy Yankee
<input type="checkbox"/>	Edit Copy Delete	Ugz6XednoX2ElrbsRTB4AaABAg	Luis Fonsi - Despacito ft. Daddy Yankee
<input type="checkbox"/>	Edit Copy Delete	UgwxXGcEfq1XTFgIHdZ4AaABAg	Luis Fonsi - Despacito ft. Daddy Yankee
<input type="checkbox"/>	Edit Copy Delete	UgzCmYFAMK8Y5_cHWx4AaABAg	Luis Fonsi - Despacito ft. Daddy Yankee
<input type="checkbox"/>	Edit Copy Delete	UgwOBGutUxio5Z2MWMd4AaABAg	Luis Fonsi - Despacito ft. Daddy Yankee
<input type="checkbox"/>	Edit Copy Delete	UgysG0YRk2F8mNTplJh4AaABAg	Luis Fonsi - Despacito ft. Daddy Yankee
<input type="checkbox"/>	Edit Copy Delete	UgwVHJ2PIPLzQFFJYM94AaABAg	Luis Fonsi - Despacito ft. Daddy Yankee

Console 1 Edit Copy Delete UgwVHJ2PIPLzQFFJYM94AaABAg Luis Fonsi - Despacito ft. Daddy Yankee

Thank you for viewing my work!